

# A TWO WEIGHT INEQUALITY FOR THE HILBERT TRANSFORM ASSUMING AN ENERGY HYPOTHESIS

MICHAEL T. LACEY, ERIC T. SAWYER, AND IGNACIO URIARTE-TUERO

ABSTRACT. Let  $\sigma$  and  $\omega$  be locally finite positive Borel measures on  $\mathbb{R}$ . Subject to the pair of weights satisfying a side condition, we characterize boundedness of the Hilbert transform  $H$  from  $L^2(\sigma)$  to  $L^2(\omega)$  in terms of the  $A_2$  condition

$$\left[ \int_I \left( \frac{|I|}{|I| + |x - x_I|} \right)^2 d\omega(x) \int_I \left( \frac{|I|}{|I| + |x - x_I|} \right)^2 d\sigma(x) \right]^{\frac{1}{2}} \leq C |I|,$$

and the two testing conditions: For all intervals  $I$  in  $\mathbb{R}$

$$\begin{aligned} \int_I H(\mathbf{1}_I \sigma)(x)^2 d\omega(x) &\leq C \int_I d\sigma(x), \\ \int_I H(\mathbf{1}_I \omega)(x)^2 d\sigma(x) &\leq C \int_I d\omega(x), \end{aligned}$$

The proof uses the beautiful corona argument of Nazarov, Treil and Volberg. There is a range of side conditions, termed Energy conditions; at one endpoint, the Energy conditions are also a consequence of the testing conditions above, and at the other endpoint they are the Pivotal Conditions of Nazarov, Treil and Volberg. We detail an example which shows that the Pivotal Conditions are not necessary for boundedness of the Hilbert transform.

## CONTENTS

1. Introduction	2
2. Necessary Conditions	7
3. Grids, Haar Function, Carleson Embedding	17
4. The Good-Bad Decomposition	20
5. Main Decomposition	22
6. Energy, Stopping Intervals, Corona Decomposition	25
7. The Carleson Measure Estimates	29
8. The Paraproducts	35
9. The Remaining Estimates	42
10. A Counterexample to the Pivotal Conditions	48
References	59

---

Research supported in part by the NSF grant 0456611.

Research supported in part by NSERC.

Research supported in part by the NSF, through grant DMS-0901524.

## 1. INTRODUCTION

We provide sufficient conditions for the two weight inequality for the Hilbert transform. Indeed, subject to a side condition, a characterization of the two weight  $L^2$  inequality is given.

For a signed measure  $\omega$  on  $\mathbb{R}$  define

$$(1.1) \quad H\omega(x) \equiv \text{p.v.} \int \frac{1}{x-y} \omega(dy).$$

A *weight*  $\omega$  is a non-negative locally finite measure. For two weights  $\omega, \sigma$ , we are interested in the inequality

$$(1.2) \quad \|H(\sigma f)\|_{L^2(\omega)} \lesssim \|f\|_{L^2(\sigma)}.$$

See Definition 1.19 below for a precise definition of *p.v.* and the meaning of (1.2). The two weight problem for the Hilbert transform is to provide a real variable characterization of the pair of weights  $\omega, \sigma$  for which inequality (1.2) holds.

What should such a characterization look like? Motivated by the very successful  $A_p$  theory for the Hilbert transform in Hunt, Muckenhoupt and Wheeden [HuMuWh], one suspects that the weights should satisfy the two weight analog of the  $A_2$  condition:

$$\sup_I \frac{1}{|I|} \int_I \omega(dx) \cdot \frac{1}{|I|} \int_I \sigma(dx) < \infty.$$

As it turns out this two weight  $A_2$  condition is not sufficient. The suggestion for additional necessary conditions comes from the  $T1$  theorem of David and Journé ‘MR763911 and the two weight theorems of the second author for fractional integral operators, [Saw3]. These conditions require the following, holding uniformly over intervals  $I$ :

$$(1.3) \quad \int_I |H(\mathbf{1}_I \sigma)|^2 \omega(dx) \leq \mathcal{H}^2 \sigma(I),$$

$$(1.4) \quad \int_I |H(\mathbf{1}_I \omega)|^2 \sigma(dx) \leq (\mathcal{H}^*)^2 \omega(I).$$

Here, we are letting  $\mathcal{H}$  and  $\mathcal{H}^*$  denote the smallest constants for which these inequalities are true uniformly over all intervals  $I$ , and we write  $\sigma(I) \equiv \int_I \sigma(dx)$ .

Clearly, (1.3) is derived from applying the inequality (1.2) to indicators of intervals. One advantage of formulating the inequality (1.2) with the measure  $\sigma$  on both sides of the inequality is that duality is then easy to derive: Interchange the roles of  $\omega$  and  $\sigma$ . Thus, the condition (1.4) is also derived from (1.2). We call these ‘testing conditions’ as they are derived from simple instances of the claimed inequality. Also, we emphasize that duality in this sense is basic to the subject, and we will appeal to it repeatedly.

In a beautiful series of papers, Nazarov, Treil and Volberg have developed a sophisticated approach toward proving the sufficiency of these testing conditions combined with an improvement of the two weight  $A_2$  condition. To describe this improvement, we define this

variant of the Poisson integral for use throughout this paper. For an interval  $I$  and measure  $\omega$ ,

$$(1.5) \quad \mathbf{P}(I, \omega) \equiv \int_{\mathbb{R}} \frac{|I|}{(|I| + \text{dist}(x, I))^2} \omega(dx)$$

$$(1.6) \quad \sup_I \mathbf{P}(I, \omega) \cdot \mathbf{P}(I, \sigma) = \mathcal{A}_2^2 < \infty.$$

The last line is the improved condition of Nazarov, Treil and Volberg. We will refer to (1.6) as simply the  $A_2$  condition. F. Nazarov has shown that even this strengthened  $A_2$  condition is not sufficient for the two weight inequality (1.2) - see e.g. Theorem 2.1 in [NiTr].

The approach of Nazarov, Treil and Volberg involves a delicate combination of ideas: random grids (see [NTV2]), weighted Haar functions and Carleson embeddings (see [NTV3]), stopping intervals (see [Vol]) and culminates in the use of these techniques with a corona decomposition in the brilliant 2004 preprint [NTV4]. Theorem 2.2 of that paper proves the sufficiency of conditions (1.6), (1.3) and (1.4) for the two weight inequality (1.2) in the presence of two additional side conditions, the *Pivotal Conditions* given by

$$(1.7) \quad \sum_{r=1}^{\infty} \omega(I_r) \mathbf{P}(I_r, \mathbf{1}_{I_0} \sigma)^2 \leq \mathcal{P}^2 \sigma(I_0),$$

(and its dual) where the inequality is required to hold for all intervals  $I_0$ , and decompositions  $\{I_r : r \geq 1\}$  of  $I_0$  into disjoint intervals  $I_r \subsetneq I_0$ . As a result they obtain the equivalence of (1.2) with the three conditions (1.6), (1.3) and (1.4) when both weights are doubling, and also when two maximal inequalities hold. Our Theorem below contains this result as a special case.

In our approach, we replace the Pivotal Condition (1.7) by certain weaker side conditions of *energy type*. We begin with a necessary form of energy.

**Definition 1.8.** For a weight  $\omega$ , and interval  $I$ , we set

$$\mathbf{E}(I, \omega) \equiv \left[ \mathbb{E}_I^{\omega(dx)} \left[ \mathbb{E}_I^{\omega(dx')} \frac{x - x'}{|I|} \right]^2 \right]^{1/2}.$$

It is important to note that  $\mathbf{E}(I, \omega) \leq 1$ , and can be quite small, if  $\omega$  is highly concentrated inside the interval  $I$ ; in particular if  $\omega \mathbf{1}_I$  is a point mass, then  $\mathbf{E}(I, \omega) = 0$ . Note also that  $\omega(I) |I|^2 \mathbf{E}(I, \omega)^2$  is the variance of the variable  $x$ , and that we have the identity

$$\mathbf{E}(I, \omega)^2 = \frac{1}{2} \mathbb{E}_I^{\omega(dx)} \mathbb{E}_I^{\omega(dx')} \frac{(x - x')^2}{|I|^2}.$$

The following *Energy Condition* is necessary for the two weight inequality:

$$(1.9) \quad \sum_{r \geq 1} \omega(I_r) \mathbf{E}(I_r, \omega)^2 \mathbf{P}(I_r, \sigma \mathbf{1}_{I_0})^2 \leq \mathcal{E}^2 \sigma(I_0),$$

where the sum is taken over all decompositions  $I_0 = \bigcup_{r=1}^{\infty} I_r$  of the interval  $I_0$  into pairwise disjoint intervals  $\{I_r\}_{r \geq 1}$ . As  $E(I, \omega) \leq 1$ , the Energy Condition is weaker than the Pivotal Condition.

As a preliminary sufficient side condition, we consider the *geometric mean* of the pivotal and energy conditions: for  $0 \leq \epsilon \leq 2$  we say that the weight pair  $(\omega, \sigma)$  satisfies the *Hybrid Energy Condition* or simply *Hybrid Condition* provided

$$(1.10) \quad \sum_{r \geq 1} \omega(I_r) E(I_r, \omega)^\epsilon P(I_r, \sigma \mathbf{1}_{I_0})^2 \leq \mathcal{E}_\epsilon^2 \sigma(I_0),$$

where the sum is taken over all decompositions  $I_0 = \bigcup_{r=1}^{\infty} I_r$ . When  $\epsilon = 2$  this is the necessary Energy Condition and when  $\epsilon = 0$  this is the Pivotal Condition. A corollary of our main theorem is that if the weight pair  $(\omega, \sigma)$  satisfies the Hybrid Condition (1.10) and its dual for some  $\epsilon < 2$ , then the two weight inequality (1.2) is equivalent to the  $A_2$  condition (1.6) and the testing conditions (1.3) and (1.4).

Later in this paper we exhibit a weight pair  $(\omega, \sigma)$  satisfying (1.3), (1.4), (1.6) and the Hybrid Conditions for some  $\epsilon < 2$ , but for which the dual Pivotal Condition *fails*. In particular this shows that the Pivotal Conditions are *not* necessary for the two weight inequality (1.2).

**1.1. An optimal condition.** Now we describe an optimal—for the method of proof—sufficient side condition. First it is convenient to introduce two functionals of pairs of sets that arise.

**Definition 1.11.** Fix  $0 \leq \epsilon < 2$ . We define the functionals

$$\begin{aligned} \Phi(J, E) &\equiv \omega(J) E(J, \omega)^2 P(J, \mathbf{1}_E \sigma)^2, \\ \Psi(J, E) &\equiv \omega(J) E(J, \omega)^\epsilon P(J, \mathbf{1}_E \sigma)^2. \end{aligned}$$

Note that  $\Phi(I_r, I_0)$  appears in the sum on the left side of the Energy Condition (1.9), and  $\Phi(J, \widehat{I} \setminus I')$  appears again on the right side of the dual Energy Estimate (6.7) below. The larger functional  $\Psi(I_r, I_0)$  appears in the sum on the left side of the Hybrid Condition (1.10).

It turns out that one can replace  $\Psi$  in the proof below with any functional, subject to three properties holding. We now describe the three properties required of the functional  $\Psi$ .

Set  $e(I) \equiv \{a, b, \frac{a+b}{2}\}$  to be the set consisting of the endpoints and midpoint of an interval  $I = [a, b]$ . We say that a subpartition  $\{J_r\}$  of  $I$  is  $\epsilon$ -good if

$$(1.13) \quad \text{dist}(J_r, e(I)) > \frac{1}{2} |J_r|^\epsilon |I|^{1-\epsilon}.$$

For  $\gamma > 0$  and  $\varepsilon > 0$ , and for all pairs of intervals  $I_0 \subset \widehat{I}$  in  $\mathcal{D}^\sigma$  we require

$$(1.14) \quad \begin{cases} \Psi(I_0, I_0) \leq \mathcal{F}_{\gamma, \varepsilon}^2 \sigma(I_0), \\ \sum_{r \geq 1} \Psi(I_r, I_0) \leq \mathcal{F}_{\gamma, \varepsilon}^2 \sigma(I_0), & \text{for all subpartitions } \{I_r\} \text{ of } I_0, \\ \sum_{r \geq 1} \Phi(J_r, \widehat{I} \setminus I_0) \leq \sup_{r \geq 1} \left( \frac{|J_r|}{|I_0|} \right)^\gamma \Psi(I_0, \widehat{I}) \\ & \text{for all } \varepsilon\text{-good subpartitions } \{J_r\} \text{ of } I_0. \end{cases}$$

**Note:** It is important to note that the second line requires us to test over *all* subpartitions  $\{I_r\}$  of  $I_0$ . In the third line we need only test over the  $\varepsilon$ -good subpartitions, but must include *differences*  $\widehat{I} \setminus I_0$  of intervals in the argument of  $\Phi$  on the left side.

When  $\Psi$  is given by (1.12), the first line in (1.14) is the usual  $A_2$  condition, the second line is the Hybrid Condition (1.10) with  $\epsilon = \gamma$ , and the third line is proved in Lemma 2.19 below.

For fixed  $\gamma, \varepsilon > 0$ , there is a *smallest* functional  $\Psi_{\gamma, \varepsilon}$  satisfying the third line in (1.14), namely

$$(1.15) \quad \Psi_{\gamma, \varepsilon}(I, E) \equiv \sup_{I \supset \bigcup_{s \geq 1} J_s} \left[ \inf_{s \geq 1} \left( \frac{|I|}{|J_s|} \right)^\gamma \right] \sum_{s \geq 1} \Phi(J_s, E),$$

where the supremum is taken over all  $\varepsilon$ -good subpartitions  $\{J_s\}$  of the interval  $I$ . Note that  $\Psi_{\gamma, \varepsilon}(I, E)$  becomes smaller as either  $\gamma$  or  $\varepsilon$  becomes smaller, and also as  $E$  becomes smaller. The functional  $\Psi_{\gamma, \varepsilon}$  also satisfies the first line as we see by taking  $E = I_0$  and the trivial decomposition  $I_1 = I_0$ . Then the second line in (1.14) becomes

$$(1.16) \quad \sum_{r \geq 1} \Psi_{\gamma, \varepsilon}(I_r, I_0) \leq \mathcal{F}_{\gamma, \varepsilon}^2 \sigma(I_0), \quad \text{for all subpartitions } \{I_r\} \text{ of } I_0.$$

This condition (1.16), which we call the *Energy Hypothesis*, thus represents the optimal side condition that can be used, along with its dual version, with the methods of this paper (all three lines in (1.14) hold and the third line is optimal). From Lemma 2.19 and the optimal property of  $\Psi_{\gamma, \varepsilon}$ , we see that the Hybrid Condition (1.10) implies the Energy Hypothesis (1.16) with  $\gamma = 2 - \epsilon - 2\varepsilon > 0$ .

**Theorem 1.17.** *Suppose that  $\omega$  and  $\sigma$  are locally finite positive Borel measures on the real line having no point masses in common, namely  $\omega(\{x\})\sigma(\{x\}) = 0$  for all  $x \in \mathbb{R}$ . Suppose in addition that for some  $\gamma > 0$ , and  $0 < \varepsilon < 1$  we have both Energy Hypothesis constants  $\mathcal{F}_{\gamma, \varepsilon}$  and  $\mathcal{F}_{\gamma, \varepsilon}^*$  finite. Then the two weight inequality (1.2) holds if and only if*

- the pair of weights satisfies the  $A_2$  condition (1.6);
- the testing conditions (1.3) and (1.4) both hold.

**Remark 1.18.** *The reader can easily check that Theorem 1.17 holds if the infimum  $\inf_{s \geq 1} \left( \frac{|I|}{|J_s|} \right)^\gamma$  in (1.15) is replaced by  $\inf_{s \geq 1} \eta \left( \frac{|I_r|}{|J_{r,s'}|} \right)$  for a suitable Dini function  $\eta$  on  $[1, \infty)$ .*

The quantitative estimate we give for the norm of the Hilbert transform is given in (5.1). Consider now the conjecture of Volberg [Vol] that the two weight inequality holds if and only if the  $A_2$  and testing conditions hold. Since the Energy Condition (1.9) is actually a consequence of the  $A_2$  and testing condition (1.3), Volberg's conjecture would be proved if we could take  $\gamma = 0$  in Theorem 1.17. (That we can take  $\varepsilon > 0$  follows from the general techniques of §4.) There are subtle obstacles to overcome in order to achieve such a characterization.

We will follow the beautiful approach of Nazarov, Treil and Volberg using random grids, stopping intervals and corona decompositions. Energy enters into the argument at those parts based upon the smoothness of the kernel, see the Energy Lemma, especially (6.6) below. Much of the argument we use appears in Chapters 17-22 of the CBMS book by Volberg [Vol], with the final touches in the preprint of Nazarov, Treil and Volberg [NTV4]. In order to make this complicated proof self-contained, we reproduce these critical ideas in our sufficiency proof below.

In §10, we exhibit a pair of weights which satisfy the two-weight inequality, as they fall within the scope of our Main Theorem, yet they do *not* satisfy the Pivotal Condition of Nazarov-Treil-Volberg.

The main novelty of this paper is that (1) the energy condition is *necessary* for the two weight testing conditions, (2) the Energy Hypothesis can be inserted into the approach of [NTV4], and (3) that the Pivotal Conditions are not necessary for the two-weight inequality.

The integral defining  $H(\sigma f)$  in (1.2) is not in general absolutely convergent, and we must introduce appropriate truncations. The following canonical construction from [Vol] serves our purposes here.

**Definition 1.19.** *Let  $\zeta$  be a fixed smooth nondecreasing function on the real line satisfying*

$$\zeta(t) = 0 \text{ for } t \leq \frac{1}{2} \text{ and } \zeta(t) = 1 \text{ for } t \geq 1.$$

*Given  $\varepsilon > 0$ , set  $\zeta_\varepsilon(t) = \zeta\left(\frac{t}{\varepsilon}\right)$  and define the smoothly truncated operator  $T_\varepsilon$  by the absolutely convergent integral*

$$T_\varepsilon f(x) = \int \frac{1}{y-x} \zeta_\varepsilon(|x-y|) f(y) d\sigma(y), \quad f \in L^2(\sigma) \text{ with compact support.}$$

*We say (1.2) holds if the inequality there holds for all compactly supported  $f$  with  $T_\varepsilon$  in place of  $T$ , uniformly in  $\varepsilon > 0$ .*

One easily verifies that all of the necessary conditions derived below can be achieved using this definition provided  $\omega$  and  $\sigma$  have no point masses in common (note that if  $\omega = \sigma =$

$\delta_x$ , then (1.2) holds trivially with this definition while (1.6) fails). Moreover, the kernels  $\frac{1}{y-x}\zeta_\varepsilon(|x-y|)$  of  $T_\varepsilon$  are uniformly standard Calderón-Zygmund kernels, and thus all of the sufficiency arguments below hold as well using this definition. In the sequel we will suppress the use of  $T_\varepsilon$  and simply write  $T$ .

## 2. NECESSARY CONDITIONS

In this section, we collect some conditions which follow either from the assumed norm inequality or the testing conditions. These are the  $A_2$  condition, a weak-boundedness condition, and the Energy Condition. The principal novelty is the Energy Condition.

**2.1. The Necessity of the  $A_2$  Condition.** In this section we will give a new proof of this known fact due to F. Nazarov:

**Proposition 2.1.** *Assuming the norm inequality (1.2), we have the  $A_2$  condition (1.6). Qualitatively,*

$$\mathcal{N} \equiv \|H(\cdot\sigma)\|_{L^2(\sigma) \rightarrow L^2(\omega)} \gtrsim \mathcal{A}_2$$

The analogue of this inequality in the unit disk was proved for the conjugate operator in [NTV4] and [Vol] Chapter 16. We provide a real-variable proof here.

*Proof.* Fix an interval  $I$  and for  $a \in \mathbb{R}$  and  $r > 0$  let

$$\begin{aligned} s_I(x) &= \frac{|I|}{|I| + |x - x_I|}, \\ f_{a,r}(y) &= \mathbf{1}_{(a-r,a)}(y) s_I(y), \end{aligned}$$

where  $x_I$  is the center of the interval  $I$ . For  $y < x$  we have

$$\begin{aligned} |I|(x-y) &= |I|(x-x_I) + |I|(x_I-y) \\ &\lesssim (|I| + |x-x_I|)(|I| + |x_I-y|), \end{aligned}$$

and so

$$\frac{1}{x-y} \geq |I|^{-1} s_I(x) s_I(y), \quad y < x.$$

Thus for  $x > a$  we obtain that

$$\begin{aligned} H(f_{a,r}\sigma)(x) &= \int_{a-r}^a \frac{1}{x-y} s_I(y) d\sigma(y) \\ &\geq |I|^{-1} s_I(x) \int_{a-r}^a s_I(y)^2 d\sigma(y). \end{aligned}$$

Applying our assumed two weight inequality (1.2) in the sense of Definition 1.19, and then letting  $\varepsilon > 0$  there go to 0, we see that

$$\begin{aligned} & |I|^{-2} \int_a^\infty s_I(x)^2 \left( \int_{a-r}^a s_I(y)^2 d\sigma(y) \right)^2 d\omega(x) \\ & \leq \|H(\sigma f_{a,r})\|_{L^2(\omega)}^2 \lesssim \mathcal{N}^2 \|f_{a,r}\|_{L^2(\sigma)}^2 = \mathcal{N}^2 \int_{a-r}^a s_I(y)^2 d\sigma(y). \end{aligned}$$

Rearranging the last inequality, we obtain

$$|I|^{-2} \int_a^\infty s_I(x)^2 d\omega(x) \int_{a-r}^a s_I(y)^2 d\sigma(y) \lesssim \mathcal{N}^2,$$

and upon letting  $r \rightarrow \infty$ , and taking a square root,

$$(2.2) \quad \left( \int_a^\infty s_I(x)^2 d\omega(x) \int_{-\infty}^a s_I(y)^2 d\sigma(y) \right)^{\frac{1}{2}} \lesssim \mathcal{N} |I|.$$

The ranges of integration are complementary half-lines, and clearly we can reverse the role of the two weights above.

Choose  $a \in \mathbb{R}$  which evenly divides the  $L^2(\sigma)$ -norm of  $s_I$  in this sense:

$$(2.3) \quad \int_{-\infty}^a s_I(y)^2 d\sigma(y) = \int_a^\infty s_I(y)^2 d\sigma(y) = \frac{1}{2} \int_{-\infty}^\infty s_I(y)^2 d\sigma(y),$$

and conclude that

$$\begin{aligned} & \int_{-\infty}^\infty s_I(x)^2 d\omega(x) \int_{-\infty}^\infty s_I(y)^2 d\sigma(y) = \int_{-\infty}^a s_I(x)^2 d\omega(x) \int_{-\infty}^\infty s_I(y)^2 d\sigma(y) \\ & \quad + \int_a^\infty s_I(x)^2 d\omega(x) \int_{-\infty}^\infty s_I(y)^2 d\sigma(y) \\ (2.4) \quad & \leq 2 \int_{-\infty}^a s_I(x)^2 d\omega(x) \int_a^\infty s_I(y)^2 d\sigma(y) \\ & \quad + 2 \int_a^\infty s_I(x)^2 d\omega(x) \int_{-\infty}^a s_I(y)^2 d\sigma(y) \\ & \lesssim \mathcal{N}^2 |I|^2. \end{aligned}$$

Dividing through by  $|I|^2$ , and forming the supremum over  $I$  concludes the proof in the case where we can choose  $a$  as in (2.3).

We now consider the case where a point masses in  $\sigma$  prevents (2.3) from holding. If we replace  $a$  by  $a + \varepsilon$  in (2.2), and then let  $\varepsilon \rightarrow 0$  this gives

$$\int_{(a,\infty)} s_I(x)^2 d\omega(x) \int_{(-\infty,a]} s_I(y)^2 d\sigma(y) \lesssim \mathcal{N}^2 |I|^2.$$



The ranges of integration are complementary half-lines, and clearly we can reverse the role of the open and closed half-lines, as well as the role of the two weights, resulting in four such inequalities altogether.

Now choose  $a \in \mathbb{R}$  to be the largest number satisfying

$$\int_{(-\infty, a)} s_I(y)^2 d\sigma(y) \leq \frac{1}{2} \int_{-\infty}^{\infty} s_I(y)^2 d\sigma(y).$$

Of course it may happen that *strict* inequality occurs in (2.1) due to a point mass in  $\sigma$  at the point  $a$ . In the event that this point mass is missing or relatively small, i.e.

$$\sigma(\{a\}) s_I(a)^2 \leq \frac{1}{2} A,$$

where  $A = \int_{-\infty}^{\infty} s_I(y)^2 d\sigma(y)$ , we can conclude that at least one of the integrals  $\int_{(-\infty, a)} s_I(y)^2 d\sigma(y)$  or  $\int_{(a, \infty)} s_I(y)^2 d\sigma(y)$  is at least  $\frac{1}{4}A$ . Suppose that the first integral  $\int_{(-\infty, a)} s_I(y)^2 d\sigma(y)$  is at least  $\frac{1}{4}A$ , and moreover is the smaller of the two if both are at least  $\frac{1}{4}A$ . Then we also have  $\int_{[a, \infty)} s_I(y)^2 d\sigma(y) \geq \frac{1}{4}A$ , where we have included the point mass at  $a$  in the integral on the left. We can now repeat the argument of (2.4) to conclude this case.

It remains to consider the case that the point mass at  $a$  is a relatively large proportion of the Poisson integral, i.e.

$$\sigma(\{a\}) s_I(a)^2 > \frac{1}{2} \int_{-\infty}^{\infty} s_I(y)^2 d\sigma(y).$$

But then, consider the two universal inequalities

$$\begin{aligned} \int_{(a, \infty)} s_I(x)^2 d\omega(x) \int_{(-\infty, a]} s_I(y)^2 d\sigma(y) &\lesssim \mathcal{N}^2 |I|^2, \\ \int_{(-\infty, a)} s_I(x)^2 d\omega(x) \int_{[a, \infty)} s_I(y)^2 d\sigma(y) &\lesssim \mathcal{N}^2 |I|^2. \end{aligned}$$

Both integrals against  $\sigma$  include the point mass at  $a$ , hence they exceed  $\frac{1}{2} \int_{-\infty}^{\infty} s_I(y)^2 d\sigma(y)$ . It is our hypothesis that  $\omega$  and  $\sigma$  do not have common point masses, so we conclude the  $A_2$  condition in this case.

□

**Remark 2.5.** *Preliminary results in this direction were obtained by Muckenhoupt and Wheeden, and in the setting of fractional integrals by Gabidzashvili and Kokilashvili, and here we follow the argument proving (1.9) in Sawyer and Wheeden [SaWh], where ‘two-tailed’ inequalities, like those in the  $A_2$  condition (1.6), originated in the fractional integral setting.*

A somewhat different approach to this for the conjugate operator in the disk uses conformal invariance and appears in [NTV4], and provides the first instance of a strengthened  $A_2$  condition being proved necessary for a two weight inequality for a singular integral.

**Remark 2.6.** In the proof of the sufficient direction of the Main Theorem 1.17, we only need ‘half’ of the  $A_2$  condition. Namely, we only need

$$\sup_I \frac{\omega(I)}{|I|} P(I, \sigma) < \infty,$$

along with the dual condition. This point could be of use in seeking to verify that a particular pair of weights satisfies the testing conditions.

**2.2. The Weak Boundedness Condition.** We show that a condition analogous to the weak-boundedness criteria of the David and Journé  $T1$  theorem is a consequence of the  $A_2$  condition and the two testing conditions (1.3) and (1.4).

For a constant  $C > 1$ , let  $\mathcal{W}_C$  be the best constant in the inequality

$$(2.7) \quad \left| \int_J H(\mathbf{1}_I \sigma) \omega(dx) \right| \leq \mathcal{W}_C \sigma(I)^{1/2} \omega(J)^{1/2},$$

where the inequality is uniform over all intervals  $I, J$  with  $\text{dist}(I, J) \leq |I| + |J|$  and  $C^{-1} \leq |I|/|J| \lesssim C$ . The exact value of  $C$  that we will need in the sufficient direction of our Theorem depends upon the choice of  $\varepsilon > 0$  in the Energy Hypothesis. It is therefore a constant, and we will simply write  $\mathcal{W}$  below.

**Proposition 2.8.** For  $C > 1$ , we have the inequality

$$\mathcal{W} \leq \min \{ \mathcal{H}, \mathcal{H}^* \} + C' A_2.$$

*Proof.* To see this we write

$$\int_J H(\mathbf{1}_I \sigma) d\omega = \int_{J_L} H(\mathbf{1}_I \sigma) d\omega + \int_{J_C} H(\mathbf{1}_I \sigma) d\omega + \int_{J_R} H(\mathbf{1}_I \sigma) d\omega,$$

where

$$\begin{aligned} J_L &= \{x \in J \setminus I : x \text{ lies to the left of } I\}, \\ J_C &= J \cap I, \\ J_R &= \{x \in J \setminus I : x \text{ lies to the right of } I\}. \end{aligned}$$

Now we easily have

$$\left| \int_{J_C} H(\mathbf{1}_I \sigma) d\omega \right| \lesssim \sqrt{\omega(J_C)} \left( \int_{J_C} |H(\mathbf{1}_I \sigma)|^2 d\omega \right)^{\frac{1}{2}}$$

$$\begin{aligned}
&\lesssim \sqrt{\omega(J)} \left( \int_I |H(\mathbf{1}_I \sigma)|^2 d\omega \right)^{\frac{1}{2}} \\
&\lesssim \sqrt{\omega(J)} \mathcal{H} \sqrt{\sigma(I)}.
\end{aligned}$$

The two remaining terms are each handled in the same way, so we treat only the first one  $\int_{J_L} H(\mathbf{1}_I) \sigma d\omega$ . We will use Muckenhoupt's characterization of Hardy's inequality [Muc] for weights  $\widehat{\omega}$  and  $\sigma$ : if  $B$  is the best constant in

$$(2.9) \quad \int_0^a \left( \int_0^x f \sigma \right)^2 d\widehat{\omega}(x) \leq B^2 \int_0^a |f|^2 d\sigma, \quad f \geq 0,$$

then,

$$(2.10) \quad B^2 \approx \sup_{0 < r < a} \left( \int_r^a d\widehat{\omega} \right) \left( \int_0^r d\sigma \right).$$

We will give the proof here assuming that  $\omega$  and  $\sigma$  have no point masses, as the general case is hard.

Without loss of generality we consider the extreme case  $J_L = (-a, 0)$  and  $I = (0, b)$  with  $0 < a < b$ . We decompose  $I = I_1 \cup I_2$  with  $I_1 = (0, a)$  and  $I_2 = (a, b)$ . First we note the easy estimate

$$\begin{aligned}
\left| \int_{J_L} H(\mathbf{1}_{I_2} \sigma) d\omega \right| &\lesssim \int_{-a}^0 \left( \int_a^b \frac{1}{y} d\sigma(y) \right) d\omega(x) = \omega(J_L) \int_a^b \frac{1}{y} d\sigma(y) \\
&\lesssim \omega(J_L) \sqrt{\sigma(I_2)} \left( \int_a^b \frac{1}{y^2} d\sigma(y) \right)^{\frac{1}{2}} \\
&\lesssim \sqrt{\omega(J_L) \sigma(I_2)} \sqrt{\frac{\omega(J_L)}{a} \mathbf{P}(I_1, \sigma)} \leq 2\mathcal{A}_2 \sqrt{\omega(J_L) \sigma(I_2)},
\end{aligned}$$

since  $J_L$  and  $I_1$  are touching intervals of equal length  $a$ . Then we use (2.10) for the other term:

$$\begin{aligned}
\left| \int_{J_L} H(\mathbf{1}_{I_1} \sigma) d\omega \right| &= \iint_{(-a, 0) \times (0, a)} \mathbf{1}_{\{-x > y\}} \frac{1}{y - x} d\sigma(y) d\omega(x) \\
&\quad + \iint_{(-a, 0) \times (0, a)} \mathbf{1}_{\{-x < y\}} \frac{1}{y - x} d\sigma(y) d\omega(x) \\
&= I + II.
\end{aligned}$$

These two terms are symmetric in  $\omega$  and  $\sigma$  so we consider only the first one  $I$ . We have letting  $z = -x$  and  $d\widetilde{\omega}(z) = d\omega(-z)$  and  $d\widehat{\omega}(z) = \frac{1}{z^2} d\widetilde{\omega}(z)$ ,

$$I = \int_0^a \int_0^z \frac{1}{y + z} d\sigma(y) d\omega(-z) \leq \int_0^a \frac{1}{z} \int_0^z d\sigma d\widetilde{\omega}(z)$$

$$\begin{aligned}
&\lesssim \left[ \int_0^a d\tilde{\omega} \int_0^a \left( \int_0^z d\sigma \right)^2 d\hat{\omega}(z) \right]^{\frac{1}{2}} \\
&\lesssim B \left[ \int_{-a}^0 d\omega \times \int_0^a d\sigma \right]^{\frac{1}{2}} = B \sqrt{\omega(J_L)\sigma(I_1)},
\end{aligned}$$

upon using (2.9) with  $f \equiv 1$ . Finally we obtain  $B \lesssim \mathcal{A}_2$  from (2.10) and

$$\begin{aligned}
\int_r^a d\hat{\omega} \int_0^r d\sigma &= \int_r^a r d\hat{\omega} \times \frac{1}{r} \int_0^r d\sigma \\
&\leq \mathbf{P}((0, r), \tilde{\omega}) \times \frac{1}{r} \int_{(0, r)} d\sigma \\
&= \mathbf{P}((-r, 0), \omega) \times \frac{1}{r} \int_{(0, r)} d\sigma \lesssim \mathcal{A}_2^2.
\end{aligned}$$

Thus we have proved  $\mathcal{W} \leq \mathcal{H} + C\mathcal{A}_2$ . We obtain  $\mathcal{W} \leq \mathcal{H}^* + C\mathcal{A}_2$  by applying the above reasoning to

$$\int_J H(\mathbf{1}_I \sigma) d\omega = \int_I H(\mathbf{1}_J \omega) d\sigma.$$

□

**2.3. The Energy Condition.** We show here that the Energy Conditions are implied by the  $\mathcal{A}_2$  and testing conditions.

**Proposition 2.11.** *We have the inequality  $\mathcal{E} \lesssim \mathcal{A}_2 + \mathcal{H}$ , and similarly for  $\mathcal{E}^*$ .*

The Energy Hypotheses with  $\gamma > 0$  are the essential tools in organizing the sufficient proof. The proof begins with this Lemma.

**Lemma 2.12.** *For any interval  $I$  and any positive measure  $\nu$  supported in  $\mathbb{R} \setminus I$ , we have*

$$(2.13) \quad \mathbf{P}(I; \nu) \leq 2|I| \inf_{x, y \in I} \frac{H\nu(x) - H\nu(y)}{x - y},$$

For specificity, in this section, we are re-defining the Poisson integral to be

$$(2.14) \quad \mathbf{P}(I; \nu) \equiv \frac{\nu(I)}{|I|} + \frac{|I|}{2} \int_{\mathbb{R} \setminus I} \frac{1}{|z - z_I|^2} \nu(dz),$$

with  $z_I$  the center of  $I$ . Note that this definition of  $\mathbf{P}(I; \nu)$  is comparable to that in (1.5). Note that  $H(\mathbf{1}_{I^c} \nu)$  is increasing on  $I$  when  $\nu$  is positive, so that the infimum in (2.13) is nonnegative.

*Proof.* To see (2.13), we suppose without loss of generality that  $I = (-a, a)$ , and a calculation then shows that for  $-a \leq x < y \leq a$ ,

$$\begin{aligned} H\nu(y) - H\nu(x) &= \int_{\mathbb{R} \setminus I} \left\{ \frac{1}{z-y} - \frac{1}{z-x} \right\} \nu(dz) \\ &= (y-x) \int_{\mathbb{R} \setminus I} \frac{1}{(z-y)(z-x)} \nu(dz) \\ &\geq \frac{1}{4} (y-x) \int_{\mathbb{R} \setminus I} \frac{1}{z^2} \nu(dz) \end{aligned}$$

since  $(z-y)(z-x)$  is positive and satisfies

$$\frac{1}{(z-y)(z-x)} \geq \frac{1}{4z^2}$$

on each interval  $(-\infty, -a)$  and  $(a, \infty)$  in  $\mathbb{R} \setminus I$  when  $-a \leq x < y \leq a$ . Thus we have from (2.14), and the assumption about the support of  $\nu$ ,

$$\begin{aligned} \mathbf{P}(I; \nu) &= \frac{|I|}{2} \int_{\mathbb{R} \setminus I} \frac{1}{z^2} \nu(dz) \\ &\lesssim 2|I| \inf_{x, y \in I} \frac{H\nu(y) - H\nu(x)}{y-x}. \end{aligned}$$

□

*Proof of Proposition 2.11.* We recall the energy condition in (1.9). Fix an interval  $I_0$ , and pairwise disjoint strict subintervals  $\{I_r : r \geq 1\}$ . Let  $\{I_s : s \geq 1\}$  be pairwise disjoint subintervals of  $I_r$ .

We apply (2.13), so that for  $x, y \in I_r$ , we have

$$\frac{|y-x|}{|I_r|} \mathbf{P}(I_r; \mathbf{1}_{I_0} \sigma) \lesssim \frac{|y-x|}{|I_r|} \frac{\sigma(I_r)}{|I_r|} + |H(\mathbf{1}_{I_0 \cap I_r^c} \sigma)(y) - H(\mathbf{1}_{I_0 \cap I_r^c} \sigma)(x)|.$$

Let us for the moment assume that the second term on the right is dominant. Squaring the inequality above, averaging with respect to the measure  $\omega$  in both  $x$  and  $y$ , we obtain

$$\begin{aligned} \mathbf{E}(I_r, \omega)^2 \mathbf{P}(I_r; \mathbf{1}_{I_0} \sigma)^2 &\leq \mathbb{E}_{I_r}^{\omega(dx)} \mathbb{E}_{I_r}^{\omega(dy)} \left( \frac{|y-x|}{|I_r|} \right)^2 \mathbf{P}(I_r; \mathbf{1}_{I_0} \sigma)^2 \\ &\lesssim \mathbb{E}_{I_r}^{\omega(dx)} \mathbb{E}_{I_r}^{\omega(dy)} |H(\mathbf{1}_{I_0 \cap I_r^c} \sigma)(y) - H(\mathbf{1}_{I_0 \cap I_r^c} \sigma)(x)|^2 \\ &\leq 4 \mathbb{E}_{I_r}^{\omega(dx)} |H(\mathbf{1}_{I_0 \cap I_r^c} \sigma)|^2. \end{aligned}$$

Multiply the last inequality by  $\omega(I_r)$  and sum in  $r$  to get

$$\begin{aligned} \sum_{r \geq 1} \omega(I_r) \mathbb{E}(I_r, \omega)^2 \mathbf{P}(I_r; \mathbf{1}_{I_0} \sigma)^2 &\lesssim \sum_r \int_{I_r} |H(\mathbf{1}_{I_0 \cap I_r^c} \sigma)|^2 d\omega \\ &\lesssim \int_{I_0} |H(\mathbf{1}_{I_0} \sigma)|^2 d\omega + C \sum_r \int_{I_r} |H(\mathbf{1}_{I_r} \sigma)|^2 d\omega \\ &\leq 2\mathcal{H}^2 \sigma(I_0) \end{aligned}$$

by the testing condition (1.3) applied to both  $I_0$  and  $I_r$ .

Returning to (2.15), it remains to consider the case where the first term on the right is dominant. By the same reasoning, we arrive at

$$\begin{aligned} \mathbb{E}(I_r, \omega)^2 \mathbf{P}(I_r; \mathbf{1}_{I_0} \sigma)^2 &\leq \mathbb{E}_{I_r}^{\omega(dx)} \mathbb{E}_{I_r}^{\omega(dy)} \frac{|y - x|^2}{|I_r|^2} \frac{\sigma(I_r)^2}{|I_r|^2} \\ &\leq \frac{\sigma(I_r)^2}{|I_r|^2}. \end{aligned}$$

Multiply the last inequality by  $\omega(I_r)$  and sum in  $r$  to get

$$\sum_{r=1}^{\infty} \frac{\sigma(I_r)^2}{|I_r|^2} \omega(I_r) \leq \mathcal{A}_2^2 \sum_{r=1}^{\infty} \sigma(I_r) \leq \mathcal{A}_2^2 \sigma(I_0).$$

□

**Remark 2.15.** We refer to  $\mathbb{E}(I, \omega)$  as the energy functional because in dimension  $n \geq 3$  the integral

$$\int_I \int_I |x - x'|^{2-n} d\omega(x) d\omega(x')$$

represents the energy required to compress charge from infinity to a distribution  $\omega$  on  $I$ , assuming a repulsive inverse square law force. When  $n = 1$ , the force is attractive and the integral

$$\int_I \int_I |x - x'| d\omega(x) d\omega(x')$$

represents the energy required to disperse charge from a point to a distribution  $\omega$  on  $I$ .

**2.4. The Hybrid Condition.** We begin with a monotonicity property of energy, and then apply it to show that the Hybrid Condition implies the Energy Hypothesis. This Lemma helps clarify the role of the Hybrid Conditions.

**Lemma 2.16.** *Fix  $0 \leq \epsilon \leq 2$ . Let  $I_0$  be an interval, and  $\{I_r : r \geq 1\}$  a partition of  $I_0$ . We have the inequalities for  $0 < \varepsilon < 1 - \frac{\epsilon}{2}$ :*

$$\sum_{r \geq 1} \omega(I_r) |I_r|^\epsilon \mathbb{E}(I_r; \omega)^\epsilon \leq \omega(I_0) |I_0|^\epsilon \mathbb{E}(I_0; \omega)^\epsilon.$$

$$\sum_{r \geq 1} \omega(I_r) |I_r|^{2-2\varepsilon} \mathbb{E}(I_r; \omega)^\epsilon \leq \sup_{r \geq 1} \left( \frac{|I_r|}{|I_0|} \right)^{2-2\varepsilon-\epsilon} \omega(I_0) |I_0|^{2-2\varepsilon} \mathbb{E}(I_0; \omega)^\epsilon.$$

The second inequality is obvious given the first; as it turns out this is the basic fact used to exploit the Hybrid Conditions for  $0 \leq \epsilon < 2$ , so we have stated it explicitly.

*Proof.* The inequality is obvious for  $\epsilon = 0$ . We prove it for  $\epsilon = 2$ . This is rather clear if we make the definition

$$\text{Var}_I^\omega \equiv \omega(I) \mathbb{E}_I^\omega(x - \mathbb{E}_I^\omega x)^2.$$

Then, we have  $\omega(I) |I|^2 \mathbb{E}(I; \omega)^2 = \text{Var}_I^\omega$ .

Second, variation of a random variable  $Z$  is the squared  $L^2$ -distance of  $Z$  from the linear space of constants. And  $\omega(I_0) |I_0|^2 \mathbb{E}(I_0; \omega)^2$  admits a transparent reformulation in this language: The random variable is  $x$  and the probability measure is normalized  $\omega$  measure. Moreover,

$$\sum_{r \geq 1} \omega(I_r) |I_r|^2 \mathbb{E}(I_r; \omega)^2$$

is the squared  $L^2$ -distance of  $x$  to the space of functions piecewise constant on the intervals of the partition  $\{I_r : r \geq 1\}$ . Hence, the inequality above is immediate.

For the case of  $0 < \epsilon < 2$ , we apply Hölder's inequality and appeal to the case of  $\epsilon = 2$ .

$$\begin{aligned} \left( \sum_{r \geq 1} \omega(I_r) |I_r|^\epsilon \mathbb{E}(I_r; \omega)^\epsilon \right)^{1/\epsilon} &\leq \omega(I_0)^{(2-\epsilon)/2\epsilon} \left( \sum_{r \geq 1} \omega(I_r) |I_r|^2 \mathbb{E}(I_r; \omega)^2 \right)^{1/2} \\ &\leq \omega(I_0)^{(2-\epsilon)/2\epsilon} \omega(I_0)^{1/2} |I_0| \mathbb{E}(I_0; \omega) \end{aligned}$$

which is the claimed inequality.  $\square$

Here is a Poisson inequality for good intervals that will see service both here and later in the paper.

**Lemma 2.17.** *Suppose that  $J \subset I \subset \widehat{I}$  and that  $\text{dist}(J, e(I)) > \frac{1}{2} |J|^\varepsilon |I|^{1-\varepsilon}$ . Then*

$$(2.18) \quad |J|^{2\varepsilon-2} \mathbb{P}(J, \sigma \mathbf{1}_{\widehat{I}_I})^2 \lesssim |I|^{2\varepsilon-2} \mathbb{P}(I, \sigma \mathbf{1}_{\widehat{I}_I})^2.$$

*Proof.* We have

$$\mathbf{P} \left( J, \sigma \chi_{\widehat{I} \setminus I} \right) \approx \sum_{k=0}^{\infty} 2^{-k} \frac{1}{|2^k J|} \int_{(2^k J) \cap (\widehat{I} \setminus I)} d\sigma,$$

and  $(2^k J) \cap (\widehat{I} \setminus I) \neq \emptyset$  requires

$$\text{dist} (J, e(I)) \lesssim |2^k J|.$$

By our distance assumption we must then have

$$|J|^\varepsilon |I|^{1-\varepsilon} \leq \text{dist} (J, e(I)) \lesssim 2^k |J|,$$

or

$$2^{-k} \lesssim \left( \frac{|J|}{|I|} \right)^{1-\varepsilon}.$$

Thus we have

$$\mathbf{P} \left( J, \sigma \chi_{\widehat{I} \setminus I} \right) \lesssim 2^{-k} \mathbf{P} \left( I, \sigma \chi_{\widehat{I} \setminus I} \right) \lesssim \left( \frac{|J|}{|I|} \right)^{1-\varepsilon} \mathbf{P} \left( I, \sigma \chi_{\widehat{I} \setminus I} \right),$$

which is the inequality (2.18).  $\square$

We can now obtain that the Hybrid Condition implies the Energy Hypothesis.

**Lemma 2.19.** *Let  $0 \leq \epsilon < 2$ . Then the functional*

$$\Psi(J, E) \equiv \omega(J) \mathbf{E}(J, \omega)^\epsilon \mathbf{P}(J, \mathbf{1}_E \sigma)^2$$

*satisfies the three properties in (1.14) with  $0 < \varepsilon < 1 - \frac{\epsilon}{2}$ . As a consequence, the Hybrid Condition (1.10) implies the Energy Hypothesis (1.16) with  $\gamma = 2 - 2\varepsilon - \epsilon$ .*

*Proof.* The first line in (1.14) is the usual  $A_2$  condition, and the second line is the Hybrid Condition (1.10) with  $\epsilon = \gamma$ . Thus we must show the third line:

$$\sum_{r \geq 1} \Phi \left( J_r, \widehat{I} \setminus I_0 \right) \leq \sup_{r \geq 1} \left( \frac{|J_r|}{|I_0|} \right)^\gamma \Psi \left( I_0, \widehat{I} \setminus I_0 \right),$$

for all  $\varepsilon$ -good subpartitions  $\{J_r\}$  of  $I_0$ , i.e. those satisfying (1.13). From Lemma 2.17 we have

$$|J_r|^{2\varepsilon-2} \mathbf{P}(J_r, \sigma \mathbf{1}_{\widehat{I} \setminus I_0})^2 \lesssim |I_0|^{2\varepsilon-2} \mathbf{P}(I_0, \sigma \mathbf{1}_{\widehat{I} \setminus I_0})^2.$$

Now use Lemma 2.16 and  $\mathbf{E}(J_r, \omega)^2 \leq \mathbf{E}(J_r, \omega)^\epsilon$  to obtain

$$\begin{aligned} \sum_{r \geq 1} \Phi \left( J_r, \widehat{I} \setminus I_0 \right) &= \sum_{r \geq 1} \omega(J_r) |J_r|^{2-2\varepsilon} \mathbf{E}(J_r, \omega)^2 |J_r|^{2\varepsilon-2} \mathbf{P}(J_r, \sigma \mathbf{1}_{\widehat{I} \setminus I_0})^2 \\ &\lesssim \sum_{r \geq 1} \omega(J_r) |J_r|^{2-2\varepsilon} \mathbf{E}(J_r, \omega)^\epsilon |I_0|^{2\varepsilon-2} \mathbf{P}(I_0, \sigma \mathbf{1}_{\widehat{I} \setminus I_0})^2 \end{aligned}$$



$$\begin{aligned}
&\leq \sup_{r \geq 1} \left( \frac{|J_r|}{|I_0|} \right)^{2-2\varepsilon-\epsilon} \times \omega(I_0) \mathbf{E}(I_0; \omega)^\epsilon \mathbf{P}(I_0, \sigma \mathbf{1}_{\widehat{I} \setminus I_0})^2 \\
&= \sup_{r \geq 1} \left( \frac{|J_r|}{|I_0|} \right)^\gamma \Psi(I_0, \widehat{I} \setminus I_0).
\end{aligned}$$

□

### 3. GRIDS, HAAR FUNCTION, CARLESON EMBEDDING

This section collects some standard facts which can be found e.g. in [Vol]. We call a collection of intervals  $\mathcal{G}$  a *grid* iff for all  $I, J \in \mathcal{G}$  we have  $I \cap J \in \{\emptyset, I, J\}$ . An interval  $I \in \mathcal{G}$  may have a *parent*  $I^{(1)}$ : The unique minimal interval  $J \in \mathcal{G}$  that strictly contains  $I$ . Recursively define  $I^{(j+1)} = (I^{(j)})^{(1)}$ . In the analysis of the paper, it will be necessary to distinguish the grid in question when passing to a parent. We accordingly set

$$(3.1) \quad \pi_{\mathcal{G}}^1(I) \equiv \text{The unique minimal interval } J \in \mathcal{G} \text{ that strictly contains } I.$$

Recursively set  $\pi_{\mathcal{G}}^{j+1}(I) = \pi_{\mathcal{G}}^1(\pi_{\mathcal{G}}^j(I))$ . Note that the definition of  $\pi_{\mathcal{G}}^1(I)$  makes sense even if  $I \notin \mathcal{G}$ .

A grid  $\mathcal{D}$  is *dyadic* if each interval  $I \in \mathcal{D}$  is union of  $I_-, I_+ \in \mathcal{D}$ , with  $I_-$  being the left-half of  $I$ , and likewise for  $I_+$ . We will refer to  $I_{\pm}$  as the *children* of  $I$ .

A dyadic grid  $\mathcal{D}$ , with weight  $\sigma$  admits the *Haar basis adapted to  $\sigma$  and  $\mathcal{D}$* . This basis is especially nice if the weight  $\sigma$  does not assign positive mass to any endpoint of an interval in  $\mathcal{D}$ . This can be achieved by e.g. a joint translation of the intervals in  $\mathcal{D}$ , and so it will be a standing assumption.

The Haar basis  $\{h_I^\sigma : I \in \mathcal{D}\}$  is explicitly defined to be

$$h_I^\sigma \equiv \frac{-\sigma(I_+) \mathbf{1}_{I_-} + \sigma(I_-) \mathbf{1}_{I_+}}{[\sigma(I_+)^2 \sigma(I_-) + \sigma(I_-)^2 \sigma(I_+)]^{1/2}} = \sqrt{\frac{\sigma(I_-) \sigma(I_+)}{\sigma(I)}} \left( -\frac{\mathbf{1}_{I_-}}{\sigma(I_-)} + \frac{\mathbf{1}_{I_+}}{\sigma(I_+)} \right),$$

with the convention that  $h_I^\sigma \equiv 0$  if the restriction of  $\sigma$  to either child  $I_-$  or  $I_+$  vanishes. The martingale properties of the Haar function are decisive, still at a couple of points, we have recourse to the formula

$$(3.2) \quad |\mathbb{E}_{I_\theta}^\sigma h_I^\sigma| = \sqrt{\frac{\sigma(I_{-\theta})}{\sigma(I) \sigma(I_\theta)}} \leq \sqrt{\frac{1}{\sigma(I_\theta)}}.$$

These functions are (1) pairwise orthogonal, (2) have  $\sigma$ -integral zero, (3) have  $L^2(\sigma)$ -norm either 0 or 1, and (4) form a basis for  $L^2(\sigma)$ . We also define

$$\Delta_I^\sigma f \equiv \langle f, h_I^\sigma \rangle_\sigma \cdot h_I^\sigma$$

where by  $\langle \cdot, \cdot \rangle_\sigma$  we mean the natural inner product on  $L^2(\sigma)$ . We then have the  $L^2(\sigma)$  identity

$$(3.3) \quad f = \sum_{I \in \mathcal{D}} \Delta_I^\sigma f,$$

for all  $f \in L^2(\sigma)$  that are supported in a dyadic interval  $I^0$  and satisfy  $\int_{I^0} f d\sigma = 0$ . We remark that by a simple reduction in (17.3) of [Vol], we only need (3.3) for such  $f$  in the proof of our theorem.

Note that (3.3) yields the Plancherel formula

$$(3.4) \quad \|f\|_{L^2(\sigma)}^2 = \sum_{I \in \mathcal{D}} |\langle f, h_I^\sigma \rangle_\sigma|^2, \quad f \in L^2(\sigma), \text{supp } f \subset I^0, \int_{I^0} f d\sigma = 0.$$

The following simple identities are basic as well. We have

$$(3.5) \quad \Delta_I^\sigma f = \{\mathbf{1}_{I_-} \mathbb{E}_{I_-}^\sigma f + \mathbf{1}_{I_+} \mathbb{E}_{I_+}^\sigma f\} - \mathbf{1}_I \mathbb{E}_I^\sigma f.$$

Consequently, for two intervals  $I_1 \subset I_2$ ,  $I_1, I_2 \in \mathcal{D}$ , the sum below is telescoping, so easily summable:

$$(3.6) \quad \sum_{I: I_1 \subsetneq I \subset I_2} \Delta_I^\sigma f(x) = \mathbb{E}_{I_1}^\sigma f - \mathbb{E}_{I_2}^\sigma f, \quad x \in I_1.$$

In these displays, we are using the notation

$$\mathbb{E}_I^\sigma \phi \equiv \sigma(I)^{-1} \int_I \phi \, \sigma(dx),$$

thus,  $\mathbb{E}_I^\sigma f$  is the average value of  $f$  with respect to the weight  $\sigma$  on interval  $I$ .

We turn to a brief description of paraproducts. The familiar Carleson Embedding Theorem is fundamental for the proof. The proof follows classical lines, using that the map  $f \rightarrow \mathbb{E}_I^\sigma f$  is type  $(\infty, \infty)$  and also weak type  $(1, 1)$  with respect to the measure  $\sum_{I \in \mathcal{D}} a_I \delta_I$  on  $\mathcal{D}$  by the Carleson condition (3.9).

**Theorem 3.7.** *Fix a weight  $\sigma$  and consider nonnegative constants  $\{a_I : I \in \mathcal{D}\}$ . The following two inequalities are equivalent:*

$$(3.8) \quad \sum_{I \in \mathcal{D}} a_I |\mathbb{E}_I^\sigma f|^2 \leq C_1 \|f\|_{L^2(\sigma)}^2,$$

$$(3.9) \quad \sum_{I \in \mathcal{D}: I \subset J} a_I \leq C_2 \sigma(J), \quad J \in \mathcal{D}.$$

Taking  $C_1$  and  $C_2$  to be the best constants in these inequalities, we have  $C_1 \approx C_2$  with the implied constant independent of  $\sigma$ .

There is another language commonly associated with the Carleson Embedding Theorem. For the purposes of this discussion, let  $\widehat{I} = I \times [0, |I|]$  be the square in the upper half-plane  $\mathbb{R}_+^2$  with face  $I$  on the real line, viewed as the boundary of  $\mathbb{R}_+^2$ . This is called the *box over  $I$* . And consider the linear map, a  $\sigma$ -weighted analog of the Poisson integral,

$$Af(x, t) \equiv \mathbb{E}_{(x-t/2, x+t/2)}^\sigma f.$$

Given a measure  $\mu$  on  $\mathbb{R}_+^2$ , this operator maps  $L^2(\mathbb{R}, \sigma)$  into  $L^2(\mathbb{R}_+^2, \mu)$  if and only if the measure  $\mu$  satisfies the *Carleson measure condition*

$$\mu(\widehat{I}) \lesssim \sigma(I),$$

The condition (3.9) above is a discrete analog of this condition, with the measure  $\mu$  being defined by a sum of Dirac point masses at the center of the tops of the boxes over  $I$ :

$$\mu \equiv \sum_{I \in \mathcal{D}} a_I \delta_{(c_I, |I|)}.$$

In seeking to verify the Carleson measure condition (3.9), there is a store of common reductions. A very simple one is that it suffices to test (3.9) for dyadic intervals  $J$  for which  $a_J \neq 0$ .

A slightly more complicated one is this: Let  $\mathcal{S} \subset \mathcal{D}$  be such that we have the estimate

$$(3.10) \quad \sum_{S \in \mathcal{S}: S \subset S_0} \mu(S) \lesssim C_1 \sigma(S_0), \quad S_0 \in \mathcal{S}.$$

That is, the measure  $\sigma$  has the Carleson measure property, provided one only sums intervals in  $\mathcal{S}$ . Now, suppose that  $\mu$  is a measure on  $\mathbb{R}_+^2$  such that for any  $S_0 \in \mathcal{S}$ , we have

$$(3.11) \quad \mu \left( \widehat{S}_0 \setminus \bigcup_{S \in \mathcal{S}: S \subsetneq S_0} \widehat{S} \right) \lesssim C_2 \sigma(S_0).$$

Here, we have the box over  $S_0$ , and we remove the smaller boxes. Then, we have  $\mu(\widehat{S}) \leq (C_1 + C_2) \sigma(S)$  for all  $S \in \mathcal{S}$ . We shall implicitly use this reduction.

In the two weight setting, a paraproduct would be, for example, an operator of the form

$$Tf = \sum_{I \in \mathcal{D}} \alpha_I \mathbb{E}_I^\sigma f \cdot h_I^\omega.$$

By the orthogonality of the Haar system, it follows that  $T$  maps  $L^2(\sigma)$  into  $L^2(\omega)$  if and only if the sequence of square coefficients  $\{\alpha_I^2 : I \in \mathcal{D}\}$  satisfies the condition (3.9). This is the type of argument we will be appealing to below.

## 4. THE GOOD-BAD DECOMPOSITION

Here we follow the random grid idea of Nazarov, Treil and Volberg as set out for example in Chapter 17 of [Vol]. The first step in the proof is to obtain two grids, one for each weight, that work well with each other. There are in fact many dyadic grids in  $\mathbb{R}$ . For any  $\beta = \{\beta_l\} \in \{0, 1\}^{\mathbb{Z}}$ , define the dyadic grid  $\mathbb{D}_\beta$  to be the collection of intervals

$$\mathbb{D}_\beta = \left\{ 2^n \left( [0, 1) + k + \sum_{i < n} 2^{i-n} \beta_i \right) \right\}_{n \in \mathbb{Z}, k \in \mathbb{Z}}$$

This parametrization of dyadic grids appears explicitly in [Hyt], and implicitly in [NTV2] section 9.1. Place the usual uniform probability measure  $\mathbb{P}$  on the space  $\{0, 1\}^{\mathbb{Z}}$ , explicitly

$$\mathbb{P}(\beta : \beta_l = 0) = \mathbb{P}(\beta : \beta_l = 1) = \frac{1}{2}, \quad \text{for all } l \in \mathbb{Z},$$

and then extend by independence of the  $\beta_l$ . Note that the *endpoints* and *centers* of the intervals in the grid  $\mathbb{D}_\beta$  are contained in  $\mathbb{Q}^{dy} + x_\beta$ , the dyadic rationals  $\mathbb{Q}^{dy} \equiv \{\frac{m}{2^n}\}_{m, n \in \mathbb{Z}}$  translated by  $x_\beta \equiv \sum_{i < 0} 2^i \beta_i \in [0, 1]$ . Moreover the pushforward of the probability measure  $\mathbb{P}$  under the map  $\beta \rightarrow x_\beta$  is Lebesgue measure on  $[0, 1]$ . The locally finite weights  $\omega, \sigma$  have at most countably many point masses, and it follows with probability one that  $\omega, \sigma$  do *not* charge an endpoint or center of any interval in  $\mathbb{D}_\beta$ .

For a weight  $\omega$ , we consider a random choice of dyadic grid  $\mathcal{D}^\omega$  on the probability space  $\Sigma^\omega$ , and likewise for second weight  $\sigma$ , with a random choice of dyadic grid  $\mathcal{D}^\sigma$  on the probability space  $\Sigma^\sigma$ .

**Notation 4.1.** We fix  $\varepsilon > 0$  for use throughout the remainder of the paper.

**Definition 4.2.** For a positive integer  $r$ , an interval  $J \in \mathcal{D}^\sigma$  is said to be  $r$ -bad if there is an interval  $I \in \mathcal{D}^\omega$  with  $|I| \geq 2^r |J|$ , and

$$\text{dist}(e(I), J) \leq \frac{1}{2} |J|^\varepsilon |I|^{1-\varepsilon}.$$

Here,  $e(J)$  is the set of three points consisting of the two endpoints of  $J$  and its center. (This is the set of discontinuities of  $h_J^\sigma$ .) Otherwise,  $J$  is said to be  $r$ -good. We symmetrically define  $J \in \mathcal{D}^\omega$  to be  $r$ -good.

The basic proposition here is:

**Proposition 4.3.** Fix a grid  $\mathcal{D}^\omega$  and  $J \in \mathcal{D}^\omega$ . Then  $\mathbb{P}(J \text{ is } r\text{-bad}) \leq C 2^{-\varepsilon r}$ .

*Proof.* Let  $I \in \mathcal{D}^\sigma$  with the same length as  $J$  and  $|I \cap J| \geq \frac{1}{2} |J|$ . Let  $s = \lfloor (1 - \varepsilon) r \rfloor$  and consider the  $s$ -fold ancestor  $\pi_{\mathcal{D}^\sigma}^s I$  of  $I$  in the grid  $\mathcal{D}^\sigma$ . We have

$$\text{dist}(e(\pi_{\mathcal{D}^\sigma}^s I), J) \leq 2^s |J| \leq |J|^\varepsilon |\pi_{\mathcal{D}^\sigma}^s I|^{1-\varepsilon}.$$

In order that

$$(4.4) \quad \text{dist}(e(\pi_{\mathcal{D}^\sigma}^r I), J) \leq \frac{1}{2} |J|^\varepsilon |\pi_{\mathcal{D}^\sigma}^r I|^{1-\varepsilon},$$

it would then be required that all of the further ancestors of  $I$  up to  $\pi_{\mathcal{D}^\sigma}^r I$ , namely  $\pi_{\mathcal{D}^\sigma}^{s+1} I, \dots, \pi_{\mathcal{D}^\sigma}^r I$ , *must share a common endpoint*. Indeed, if not there is  $1 \leq \ell \leq r - s$  such that

$$\begin{aligned} \text{dist}(e(\pi_{\mathcal{D}^\sigma}^r I), J) &\geq \text{dist}(e(\pi_{\mathcal{D}^\sigma}^{s+\ell} I), J) \geq \frac{1}{2} |\pi_{\mathcal{D}^\sigma}^{s+\ell} I| = 2^{s+\ell-1} |I| \\ &\geq 2^{s+1} |I| > |J|^\varepsilon 2^{(1-\varepsilon)r} |I|^{1-\varepsilon} \geq |J|^\varepsilon |\pi_{\mathcal{D}^\sigma}^r I|^{1-\varepsilon}. \end{aligned}$$

The essential point about the random construction of the grids used here is that for any interval  $K$ ,  $K$  is equally likely to be the left or right half of its parent, and the selection of parents is done independently. But sharing a common endpoint means that  $\pi_{\mathcal{D}^\sigma}^t I$  has to be the left-half, say, of  $\pi_{\mathcal{D}^\sigma}^{t+1} I$ , for all  $t = s, \dots, r - 1$ . So the probability that (4.4) holds is at most  $2^{-r+s+2} \leq 2^{-\varepsilon r+2}$ . Now by definition,  $J$  is  $r$ -bad if at least *one* of the ancestors  $\{\pi_{\mathcal{D}^\sigma}^{r+k} I\}_{k=0}^\infty$  at or beyond  $\pi_{\mathcal{D}^\sigma}^r I$  satisfies

$$(4.5) \quad \text{dist}(e(\pi_{\mathcal{D}^\sigma}^{r+k} I), J) \leq \frac{1}{2} |J|^\varepsilon |\pi_{\mathcal{D}^\sigma}^{r+k} I|^{1-\varepsilon}.$$

The argument above shows that the probability that (4.5) holds is at most  $2^{-\varepsilon(r+k)+2}$ . It follows that the probability that  $J$  is  $r$ -bad is at most

$$\sum_{k=0}^\infty 2^{-\varepsilon(r+k)+2} = 2^{-\varepsilon r+2} \sum_{k=0}^\infty 2^{-\varepsilon k} = \frac{2^{-\varepsilon r+2}}{1 - 2^{-\varepsilon}} \leq C_\varepsilon 2^{-\varepsilon r}.$$

which proves the Proposition.  $\square$

We restate the previous Proposition in a new setting. Let  $\mathcal{D}^\sigma$  be randomly selected, with parameter  $\beta$ , and  $\mathcal{D}^\omega$  with parameter  $\beta'$ . Define a projection

$$\mathbf{P}_{\text{good}}^\sigma f \equiv \sum_{I \text{ is } r\text{-good} \in \mathcal{D}^\sigma} \Delta_I^\sigma f,$$

and likewise for  $\mathbf{P}_{\text{good}}^\omega \phi$ . We define  $\mathbf{P}_{\text{bad}}^\sigma f \equiv f - \mathbf{P}_{\text{good}}^\sigma f$ . The basic Proposition is:

**Proposition 4.6.** (*Theorem 17.1 in [Vol]*) *We have the estimates*

$$\mathbb{E}_{\beta'} \|\mathbf{P}_{\text{bad}}^\sigma f\|_{L^2(\sigma)} \leq C 2^{-\frac{\varepsilon r}{2}} \|f\|_{L^2(\sigma)}.$$

*and likewise for  $\mathbf{P}_{\text{bad}}^\omega \phi$ .*

*Proof.* We have

$$\mathbb{E}_{\beta'} \|\mathbf{P}_{\text{bad}}^\sigma f\|_{L^2(\sigma)}^2 = \mathbb{E}_{\beta'} \sum_{I \text{ is } r\text{-bad}} \langle f, h_I^\sigma \rangle^2$$

$$\leq C 2^{-\varepsilon r} \sum_I \langle f, h_I^\sigma \rangle^2 = C 2^{-\varepsilon r} \|f\|_{L^2(\sigma)}^2.$$

□

From this we conclude the following: There is an absolute choice of  $r$  so that the following holds. Let  $T : L^2(\sigma) \rightarrow L^2(\omega)$  be a bounded linear operator. We then have

$$(4.7) \quad \|T\|_{L^2(\sigma) \rightarrow L^2(\omega)} \leq 2 \sup_{\|f\|_{L^2(\sigma)}=1} \sup_{\|\phi\|_{L^2(\omega)}=1} \mathbb{E}_\beta \mathbb{E}_{\beta'} |\langle T P_{\text{good}}^\sigma f, P_{\text{good}}^\omega \phi \rangle_\omega|.$$

Indeed, we can choose  $f \in L^2(\sigma)$  of norm one, and  $\phi \in L^2(\omega)$  of norm one, and we can write

$$f = P_{\text{good}}^\sigma f + P_{\text{bad}}^\sigma f$$

and similarly for  $\phi$ , so that

$$\begin{aligned} \|T\|_{L^2(\sigma) \rightarrow L^2(\omega)} &= \langle T f, \phi \rangle_\omega \\ &\leq \mathbb{E}_\beta \mathbb{E}_{\beta'} |\langle T P_{\text{good}}^\sigma f, P_{\text{good}}^\omega \phi \rangle_\omega| + \mathbb{E}_\beta \mathbb{E}_{\beta'} |\langle T P_{\text{bad}}^\sigma f, P_{\text{good}}^\omega \phi \rangle_\omega| \\ &\quad + \mathbb{E}_\beta \mathbb{E}_{\beta'} |\langle T P_{\text{good}}^\sigma f, P_{\text{bad}}^\omega \phi \rangle_\omega| + \mathbb{E}_\beta \mathbb{E}_{\beta'} |\langle T P_{\text{bad}}^\sigma f, P_{\text{bad}}^\omega \phi \rangle_\omega| \\ &\leq \mathbb{E}_\beta \mathbb{E}_{\beta'} |\langle T P_{\text{good}}^\sigma f, P_{\text{good}}^\omega \phi \rangle_\omega| + 3C \cdot 2^{-r/16} \|T\|_{L^2(\sigma) \rightarrow L^2(\omega)}. \end{aligned}$$

And this proves (4.7) for  $r$  sufficiently large depending on  $\varepsilon > 0$ .

This has the following implication for us: *It suffices to consider only  $r$ -good intervals, and prove an estimate for  $\|H(\sigma \cdot)\|_{L^2(\sigma) \rightarrow L^2(\omega)}$  that is independent of this assumption.* Accordingly, we will call  $r$ -good intervals just *good intervals* from now on.

## 5. MAIN DECOMPOSITION

Fix (large) intervals  $I^0 \in \mathcal{D}^\sigma$  and  $J^0 \in \mathcal{D}^\omega$ , and consider the following modification of the ‘good’ projections

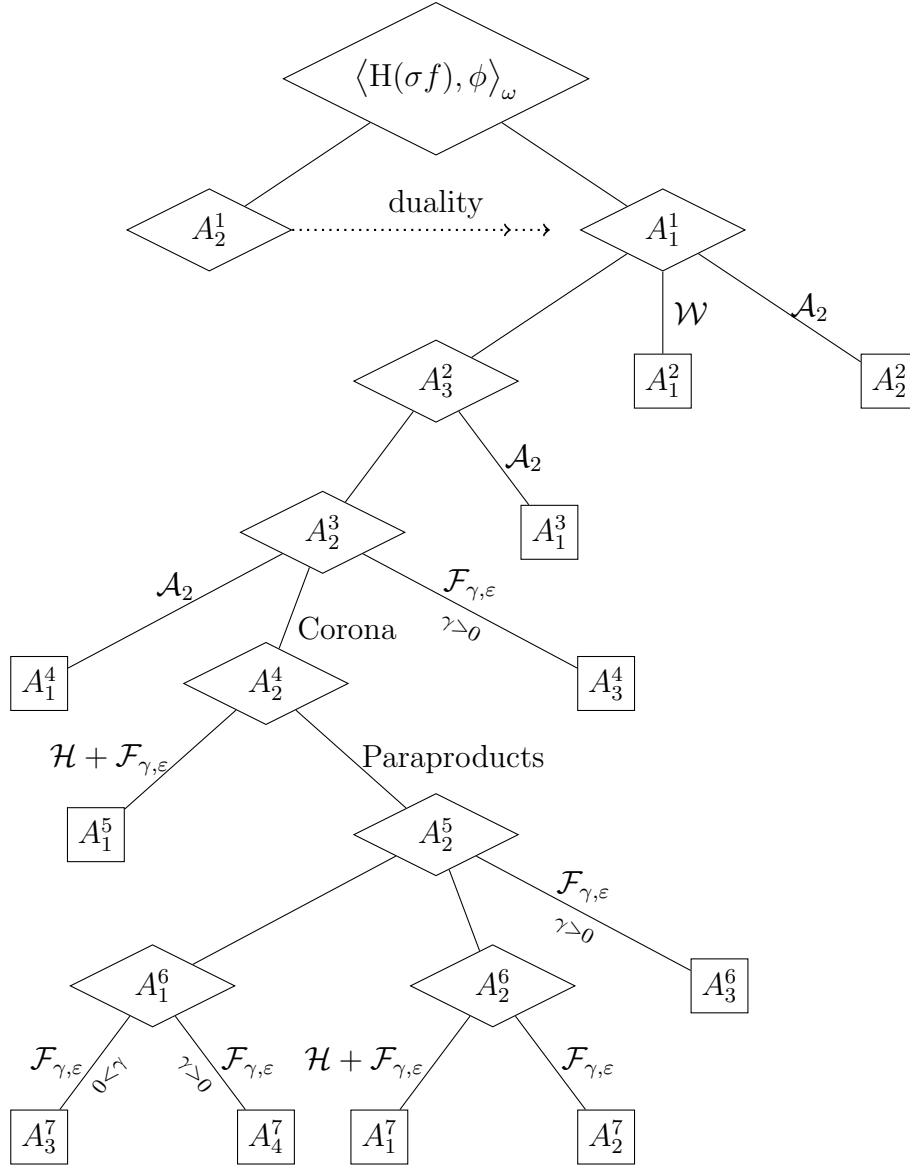
$$P_{\text{good}, I^0}^\sigma f \equiv \sum_{I \in \mathcal{D}^\sigma, I \subset I^0 \text{ and } |I| \leq 2^{-r}|I^0|, I \text{ good}} \Delta_I^\sigma f.$$

Likewise, define  $P_{\text{good}, J^0}^\omega \phi$  as above. We will prove that

$$(5.1) \quad \left| \langle H(\sigma P_{\text{good}, I^0}^\sigma f), P_{\text{good}, J^0}^\omega \phi \rangle_\omega \right| \lesssim \max\{\mathcal{A}_2, \mathcal{H}, \mathcal{H}^*, \mathcal{F}_{\gamma, \varepsilon}, \mathcal{F}_{\gamma, \varepsilon}^*\} \|f\|_{L^2(\sigma)} \|\phi\|_{L^2(\omega)}.$$

As this estimate will hold for all  $I^0, J^0$ , and all joint shifts of  $\mathcal{D}^\sigma$  and  $\mathcal{D}^\omega$  that avoid point masses at the boundary of intervals, this is sufficient to derive the Main Theorem 1.17. We also use here that the constant terms associated with the initial intervals  $I^0$  and  $J^0$  in the expansion of  $f$  and  $\phi$  respectively can be handled by the weak boundedness condition (2.8). This means we can assume the expectations  $\mathbb{E}_{I_0}^\sigma f$  and  $\mathbb{E}_{J_0}^\omega \phi$  both vanish.

We may assume that  $P_{\text{good}, I^0}^\sigma f = f$ , and likewise for  $\phi$ . **From this point forward, we will only consider good intervals  $I, J$ .** We suppress this dependence in the notation and we will clearly note the use of this hypothesis when it arises. Similarly we will only consider intervals  $I, J$  that contribute to the definition of the projections  $P_{\text{good}, I^0}^\sigma$  and  $P_{\text{good}, J^0}^\omega$ , and

FIGURE 1. The flow chart of the decomposition of the inner product  $\langle H(\sigma f), \phi \rangle_\omega$ .

suppress this fact in the notation. The role of  $I^0$  and  $J^0$  permit the recursive constructions of the *stopping intervals* in Definition 6.8 below.

Now, the inner product in (5.1) is

$$\begin{aligned} \sum_{I \in \mathcal{D}^\sigma} \sum_{J \in \mathcal{D}^\omega} \langle H(\sigma \Delta_I^\sigma f), \Delta_J^\omega \phi \rangle_\omega &= \sum_{I \in \mathcal{D}^\sigma} \sum_{J \in \mathcal{D}^\omega} \langle f, h_I^\sigma \rangle_\sigma \langle H(\sigma h_I^\sigma), h_J^\omega \rangle_\omega \langle \phi, h_J^\omega \rangle_\omega \\ &= A_1^1 + A_2^1, \end{aligned}$$

where  $\mathcal{A}_1^1 \equiv \{(I, J) \in \mathcal{D}^\sigma \times \mathcal{D}^\omega : |J| \leq |I|\}$ , and we use the notation

$$A_j^i \equiv \sum_{(I, J) \in \mathcal{A}_j^i} \langle f, h_I^\sigma \rangle_\sigma \langle H(\sigma h_I^\sigma), h_J^\omega \rangle_\omega \langle \phi, h_J^\omega \rangle_\omega .$$

The term  $A_2^1$  is the complementary sum. The sums are estimated symmetrically. Thus it suffices to prove (5.1) for the sum  $A_1^1$ . Indeed, the starred constants do not enter into this estimate, but by duality will enter into those for  $A_2^1$ , in which the roles of  $\omega$  and  $\sigma$  are reversed.

We shall follow the argument outlined in Chapters 18-22 of [Vol] by making several more decompositions, generating a number of terms  $A_j^i$ . These will be bilinear forms, but we will suppress the dependence of these forms on the functions  $f$  and  $\phi$ . In this notation, the superscript  $i$  denotes the generation of the decomposition, and we will go to seven generations. The subscript  $j$  counts the number of decompositions in a generation. To aid the reader's understanding of the argument, a flow chart of the decompositions is given in Figure 1. It contains information about the proof, which we describe here.

- The chart is read from top to bottom, with the root of the chart containing the inner product  $\langle H(\sigma f), \phi \rangle_\omega$ .
- Terms in diamonds are further decomposed, while terms in rectangles are final estimates. The edges leading into rectangles are labeled by the hypotheses used to control them,  $\mathcal{A}_2$ ,  $\mathcal{H}$ ,  $\mathcal{W}$ , or  $\mathcal{F}_{\gamma, \varepsilon}$  in the figure.
- There are three terms,  $A_3^4$ ,  $A_3^6$  and the two from  $A_1^6$  for which the Energy Hypothesis with  $\gamma > 0$  is essential. The edges leading into these terms are labeled to indicate this.
- The horizontal dotted arrow from  $A_2^1$  to  $A_1^1$ , labeled ‘duality’, indicates that  $A_2^1$  is controlled by the argument for  $A_1^1$ , after exchanging the roles of  $f$  and  $\phi$ . Accordingly, the final estimates in the dual tree will be in terms of the dual hypotheses, namely  $\mathcal{F}_{\gamma, \varepsilon}^*$  and  $\mathcal{H}^*$ .
- The edge leading into  $A_2^4$  is labeled ‘Corona’ to indicate that the Corona decomposition of §6 is used at this point. This is an important stage in the decomposition and one of the key ideas in [NTV4]. We modify it with the use of our Energy Hypothesis.
- The edge leading into  $A_2^5$  is labeled ‘Paraproducts’ as all of the estimates in the fifth and subsequent generations use paraproduct arguments to control them. See §8. We organize the written proof to pass to the paraproducts first, with the other estimates taken up second in §9.

We return to the main line of the proof of sufficiency. The collection of pairs of intervals  $\mathcal{A}_1^1$  is decomposed into the collections

$$(5.2) \quad \mathcal{A}_1^2 \equiv \{(I, J) \in \mathcal{A}_1^1 : 2^{-r}|I| \leq |J| \leq |I|, \text{dist}(I, J) \leq |I|\},$$

$$\mathcal{A}_2^2 \equiv \{(I, J) \in \mathcal{A}_1^1 : |J| \leq |I|, \text{dist}(I, J) > |I|\},$$

$$\mathcal{A}_3^2 \equiv \{(I, J) \in \mathcal{A}_1^1 : |J| < 2^{-r}|I|, \text{dist}(I, J) \leq |I|\}.$$



Using the notation of [Vol] and [NTV4], we refer to  $\mathcal{A}_1^2$  as the ‘diagonal short-range’ terms;  $\mathcal{A}_2^2$  are the ‘long-range’ terms; and  $\mathcal{A}_3^2$  are the ‘short-range’ terms.

We will show in §9.2 and §9.3, respectively,

$$(5.3) \quad |A_1^2| \lesssim \mathcal{H} \|f\|_{L^2(\sigma)} \|\phi\|_{L^2(\omega)} ,$$

$$(5.4) \quad |A_2^2| \lesssim \mathcal{A}_2 \|f\|_{L^2(\sigma)} \|\phi\|_{L^2(\omega)} .$$

These inequalities are obtained in Chapters 18 and 19 of [Vol].

The term  $\mathcal{A}_3^2$  is the important one, and will be further decomposed into

$$(5.5) \quad \mathcal{A}_1^3 \equiv \{(I, J) \in \mathcal{D}^\sigma \times \mathcal{D}^\omega : |J| < 2^{-r}|I|, I \cap J = \emptyset, \text{dist}(I, J) \leq |I|\}$$

$$(5.6) \quad \mathcal{A}_2^3 \equiv \{(I, J) \in \mathcal{D}^\sigma \times \mathcal{D}^\omega : |J| < 2^{-r}|I|, I \cap J \neq \emptyset\} .$$

The ‘mid-range’ term  $A_1^3$  will be handled by a variant of the method used on the ‘long-range’ term, along with the  $A_2$  condition. In particular, in §9.4 we prove

$$(5.7) \quad |A_1^3| \lesssim \mathcal{F}_{\gamma, \varepsilon} \|f\|_{L^2(\sigma)} \|\phi\|_{L^2(\omega)} .$$

Thus,  $A_2^3$  is the true ‘short-range’ term. It is imperative to observe that for  $(I, J) \in \mathcal{A}_2^3$ , we must have  $J \subset I$ , for otherwise we violate the fact that  $J$  is good.

## 6. ENERGY, STOPPING INTERVALS, CORONA DECOMPOSITION

Our focus is on the short-range term, as given by (5.6), and it is here that our Energy Condition (1.9) will arise in place of the Pivotal Condition in [NTV4]. This is a critical section in this proof, and it has three purposes. First, to derive the Energy Lemma, and combine it with the Energy Hypothesis. Second, to make the definition of the Corona. Third, use the Corona to obtain the next stage in the decomposition of the short-range term.

**6.1. The Energy Lemma.** As is typical in proofs that involve identification of a paraproduct, one should add and subtract cancellative terms, in order that the paraproducts become more apparent. Take a pair  $(I, J) \in \mathcal{A}_2^3$ . Thus,  $J \cap I \neq \emptyset$  and  $|J| \leq 2^{-r}|I|$ . But  $J$  is good, so that we have  $J \subset I$ , but not only that, we have

$$\text{dist}(e(I), J) \geq |J|^\varepsilon |I|^{1-\varepsilon} .$$

Let  $I_J$  be the child of  $I$  that contains  $J$ , and let  $\widehat{I}$  denote some ancestor of  $I_J$ . (The specific sequence of ancestors is to be selected below, in Definition 6.8.) We write

$$(6.1) \quad \begin{aligned} \langle H(\sigma \Delta_I^\sigma f), \Delta_J^\omega \phi \rangle_\omega &= \langle H(\mathbf{1}_{I \setminus I_J} \sigma \Delta_I^\sigma f), \Delta_J^\omega \phi \rangle_\omega + \langle H(\mathbf{1}_{I_J} \sigma \Delta_I^\sigma f), \Delta_J^\omega \phi \rangle_\omega \\ &= \langle H(\mathbf{1}_{I \setminus I_J} \sigma \Delta_I^\sigma f), \Delta_J^\omega \phi \rangle_\omega \end{aligned}$$

$$(6.2) \quad + \mathbb{E}_{I_J}^\sigma \Delta_I^\sigma f \cdot \langle H(\sigma \mathbf{1}_{\widehat{I}}), \Delta_J^\omega \phi \rangle_\omega$$

$$(6.3) \quad - \mathbb{E}_{I_J}^\sigma \Delta_I^\sigma f \cdot \langle H(\sigma \mathbf{1}_{\widehat{I} \setminus I_J}), \Delta_J^\omega \phi \rangle_\omega$$

As in [Vol] and [NTV4], we refer to the three terms in the last line as, respectively, the ‘neighbor’ term, the ‘paraproduct’ term (called ‘middle’ term in [Vol] and [NTV4]), and the ‘stopping’ term. Note that  $\Delta_I^\sigma f$  takes a single value on  $I_J$ , which is exactly  $\mathbb{E}_{I_J}^\omega \Delta_I^\sigma f$ .

Our analysis of the stopping term in (6.3) will bring forward the energy condition (1.9), and yields a more general inequality which we formulate in this Lemma. Recall that

$$\Phi(I, E) \equiv \omega(I) \mathbf{E}(I, \omega)^2 \mathbf{P}(I, \mathbf{1}_E \sigma)^2.$$

**Lemma 6.4 (Energy Lemma).** *Let  $J \subset I' \subset \widehat{I}$  be three intervals with*

$$(6.5) \quad \text{dist}(\partial I', J) \geq |J|$$

*(This follows from the good property for dyadic intervals, but we do not assume that any of these three intervals are dyadic.) Let  $\Phi_J$  be a function supported in  $J$  and with  $\omega$ -integral zero. Then we have*

$$(6.6) \quad \left| \left\langle H(\mathbf{1}_{\widehat{I} \setminus I'} \sigma), \Phi_J \right\rangle_\omega \right| \leq C \|\Phi_J\|_{L^2(\omega)} \Phi(J, \widehat{I} \setminus I')^{\frac{1}{2}}.$$

The  $L^2$  formulation in (6.6) proves useful in many estimates below, in particular in the proof of the Carleson measure estimate, Theorem 7.11. Indeed, we will apply (6.6) in the dual formulation. Namely, we have

$$(6.7) \quad \left\| H(\mathbf{1}_{\widehat{I} \setminus I'} \sigma) - \mathbb{E}_J^\omega H(\mathbf{1}_{\widehat{I} \setminus I'} \sigma) \right\|_{L^2(J, \omega)} \lesssim \Phi(J, \widehat{I} \setminus I')^{\frac{1}{2}}.$$

Note that on the left, we are subtracting off the mean value, and only testing the  $L^2(\omega)$  norm on  $J$ .

*Proof.* For  $x, x' \in J$ , and  $y \in \widehat{I} \setminus I'$ , we have the equality

$$\frac{1}{x-y} - \frac{1}{x'-y} = \frac{x-x'}{|J|} \cdot \frac{|J|}{(x-y)(x'-y)}$$

We use (6.5) to estimate the second term by

$$\frac{|J|}{(x-y)(x'-y)} \lesssim \frac{|J|}{|y - c_J|^2},$$

where  $c_J$  is the center of  $J$ .

Turning to the inner product, the fact that  $\Phi_J$  is supported on  $J$  and has  $\omega$ -mean zero permits us the usual cancellative estimate on the kernel. This familiar argument requires the selection of an auxiliary point in  $J$ , and we use the measure  $\omega$  to select it. We have

$$\left| \left\langle H(\sigma \mathbf{1}_{\widehat{I} \setminus I'}), \Phi_J \right\rangle_\omega \right| = \left| \left\langle H(\sigma \mathbf{1}_{\widehat{I} \setminus I'}) - \mathbb{E}_J^\omega H(\sigma \mathbf{1}_{\widehat{I} \setminus I'}), \Phi_J \right\rangle_\omega \right|$$

$$\begin{aligned}
&= \left| \int_J \int_{I \setminus I'} \mathbb{E}_J^{\omega(dx')} \left( \frac{1}{x-y} - \frac{1}{x'-y} \right) \Phi_J(x) \sigma(dy) \omega(dx) \right| \\
&\lesssim \int_J \mathbb{E}_J^{\omega(dx')} \frac{|x-x'|}{|J|} |\Phi_J(x)| \omega(dx) \cdot \mathbf{P}(J, \mathbf{1}_{\widehat{I} \setminus I'} \sigma) \\
&\leq \|\Phi_J\|_{L^2(\omega)} \left( \int_J \left( \mathbb{E}_J^{\omega(dx')} \frac{|x-x'|}{|J|} \right)^2 \omega(dx) \right)^{1/2} \mathbf{P}(J, \mathbf{1}_{\widehat{I} \setminus I'} \sigma) \\
&= \|\Phi_J\|_{L^2(\omega)} \Phi \left( J, \widehat{I} \setminus I' \right)^{\frac{1}{2}}.
\end{aligned}$$

This completes the proof.  $\square$

**6.2. The Corona Decomposition.** We now make two important definitions from [NTV4]: ‘stopping intervals’ and the ‘Corona Decomposition’. This is the main point of departure for our proof. But first we recall the notation introduced in (1.12), (1.15),

$$\begin{aligned}
\Phi(I, E) &\equiv \omega(I) \mathbf{E}(I, \omega)^2 \mathbf{P}(I, \mathbf{1}_E \sigma)^2, \\
\Psi_{\gamma, \varepsilon}(I, E) &\equiv \sup_{I = \bigcup_{s \geq 1} J_s} \inf_{J_s} \left[ \frac{|I|}{|J_s|} \right]^\gamma \times \sum_{s \geq 1} \Phi(J_s, E),
\end{aligned}$$

where the supremum is over all  $\varepsilon$ -good subpartitions  $\{J_s\}_{s \geq 1}$  of  $I$ , and the Energy Hypothesis (1.16)

$$\sum_{r \geq 1} \Psi_{\gamma, \varepsilon}(I_r, I_0) \leq \mathcal{F}_{\gamma, \varepsilon}^2 \sigma(I_0),$$

where  $\gamma > 0, \varepsilon > 0$  are fixed. Recall that  $\Phi$  appears in the Energy Condition (1.9) and in the dual Energy Estimate (6.7), while the larger functional  $\Psi_{\gamma, \varepsilon}$  appears in the Energy Hypothesis (1.16). The key properties required of  $\Psi$  are given in (1.14), and result in the crucial off-diagonal decay of  $\Psi$  relative to  $\Phi$  in Theorem 7.11 used to estimate term  $A_3^6$ , as well as the estimates for the term  $A_1^6$  and the stopping term  $A_3^4$  in Subsection 9.1.

**Definition 6.8.** Given any interval  $I_0$ , set  $\mathcal{S}(I_0)$  to be the maximal  $\mathcal{D}^\sigma$  strict subintervals  $S \subsetneq I_0$  such that

$$(6.9) \quad \Psi_{\gamma, \varepsilon}(S, I_0) \geq 4\mathcal{F}_{\gamma, \varepsilon}^2 \sigma(S),$$

The collection  $\mathcal{S}(I_0)$  can be empty.

We now recursively define  $\mathcal{S}_1 \equiv \{I^0\}$ , and  $\mathcal{S}_{j+1} \equiv \bigcup_{S \in \mathcal{S}_j} \mathcal{S}(S)$ . The collection  $\mathcal{S} \equiv \bigcup_{j=1}^\infty \mathcal{S}_j$  is the collection of stopping intervals. Define  $\rho : \mathcal{S} \rightarrow \mathbb{N}$  by  $\rho(S) = j$  for all  $S \in \mathcal{S}_j$ , so that  $\rho(S)$  denotes the ‘generation’ in which  $S$  occurs in the construction of  $\mathcal{S}$ .

**Remark 6.10.** *It is worth emphasizing that we will not have a uniform inequality of the following nature available to us:*

$$\Psi_{\gamma,\varepsilon}(S, I_0) \lesssim \sigma(S).$$

*In a similar, but different direction, one might be tempted to make the simpler definition of a stopping interval that it is a maximal subinterval  $S \subsetneq I_0$  for which one has*

$$\Phi(S, I_0) = \mathbf{E}(S, \omega)^2 \mathbf{P}(S, \mathbf{1}_{I_0} \sigma)^2 \omega(S) \geq 4\mathcal{E}^2 \sigma(S).$$

*This simpler condition does not permit one to fully exploit the Energy Hypothesis.*

We now define the associated Corona Decomposition.

**Definition 6.11.** *For  $S \in \mathcal{S}$ , we set  $\mathcal{P}(S)$  to be all the pairs of intervals  $(I, J)$  such that*

- (1)  $I \in \mathcal{D}^\sigma$ ,  $J \in \mathcal{D}^\omega$ ,  $J \subset I$ , and  $|J| < 2^{-r}|I|$ .
- (2)  $S$  is the  $\mathcal{S}$ -parent of  $I_J$ , the child of  $I$  that contains  $J$ .

*Note that  $\mathcal{A}_2^3 = \bigcup_{S \in \mathcal{S}} \mathcal{P}(S)$ , where  $\mathcal{A}_2^3$  is defined in (5.6). Let  $\mathcal{C}^\sigma(S)$  to be all those  $I \in \mathcal{D}^\sigma$  such that  $S$  is a minimal member of  $\mathcal{S}$  that contains a  $\mathcal{D}^\sigma$ -child of  $I$ . (A fixed interval  $I$  can be in two collections  $\mathcal{C}^\sigma(S)$ .) The definition of  $\mathcal{C}^\omega(S)$  is similar but not symmetric: all those  $J \in \mathcal{D}^\omega$  such that  $S$  is the smallest member of  $\mathcal{S}$  that contains  $J$  and satisfies  $2^r|J| < |S|$ . The collections  $\{\mathcal{C}^\sigma(S) : S \in \mathcal{S}\}$  and  $\{\mathcal{C}^\omega(S) : S \in \mathcal{S}\}$  are referred to as the Corona Decompositions. Note that  $\mathcal{S} \subset \mathcal{D}^\sigma$  and  $\mathcal{C}^\sigma(S) \subset \mathcal{D}^\sigma$  for  $S \in \mathcal{S}$  while  $\mathcal{C}^\omega(S) \subset \mathcal{D}^\omega$  for  $S \in \mathcal{S}$ .*

*We denote the associated projections by*

$$(6.12) \quad \mathbf{P}_S^\sigma f \equiv \sum_{I \in \mathcal{C}^\sigma(S)} \langle f, h_I^\sigma \rangle_\sigma h_I^\sigma,$$

*and similarly for  $\mathbf{P}_S^\omega \phi$ . Note that  $\mathbf{P}_S^\omega$  projects only on intervals  $J$  with  $|J| < 2^{-r}|S|$ .*

We have the estimate below that we will appeal to a few times.

$$(6.13) \quad \sum_{S \in \mathcal{S}} \|\mathbf{P}_S^\sigma f\|_{L^2(\sigma)}^2 \leq 2\|f\|_{L^2(\sigma)}^2$$

There is a similar inequality for  $\mathbf{P}_S^\omega$  which we will also use.

**Remark 6.14.** *In the definition of the stopping intervals, we are using the functional  $\Psi_{\gamma,\varepsilon}$  associated with the Energy Hypothesis (1.16). Thus the stopping intervals can be viewed as the enemy in verifying (1.16).*

**6.3. The Decomposition of the Short-Range Term.** To conclude this section, our estimate of  $A_2^3$  defined in (5.6) combines the splitting (6.1), (6.2), (6.3) and the Corona Decomposition. Namely, the Corona Decomposition selects the intervals  $\widehat{I}$  that appear in (6.1)—(6.3) according to the following rule. Recall that

$$\mathcal{A}_2^3 \equiv \{(I, J) \in \mathcal{D}^\sigma \times \mathcal{D}^\omega : J \subset I \text{ and } |J| < 2^{-r}|I|\}.$$

Here, we have used the fact that  $J$  is good to make the condition defining  $\mathcal{A}_2^3$  more explicit, i. e.  $J \subset I$  and  $2^r|J| < |I|$ .

**Definition 6.15.** *Given a pair  $(I, J) \in \mathcal{A}_2^3$ , choose  $\widehat{I} \in \mathcal{S}$  to be the unique stopping interval such that  $I_J \in \mathcal{C}^\sigma(\widehat{I})$ , where  $I_J$  is the child of  $I$  containing  $J$ . Equivalently,  $\widehat{I} \in \mathcal{S}$  is determined by the requirement that  $(I, J) \in \mathcal{P}(\widehat{I})$ .*

Note that if  $I_J \notin \mathcal{S}$ , then  $\widehat{I} \supset I$ , while if  $I_J \in \mathcal{S}$ , then  $\widehat{I}$  is the child of  $I$  containing  $J$ . Thus  $\widehat{I}$  is a function of the pair  $(I, J)$ . With this choice of  $\widehat{I}$  in the splitting (6.1), (6.2), (6.3) we obtain  $|A_2^3| \leq \sum_{j=1}^3 |A_j^4|$  where

$$(6.16) \quad A_1^4 \equiv \sum_{(I,J) \in \mathcal{A}_2^3} \langle H(\mathbf{1}_{I \setminus I_J} \sigma \Delta_I^\sigma f), \Delta_J^\omega \phi \rangle_\omega$$

$$(6.17) \quad A_2^4 \equiv \sum_{S \in \mathcal{S}} \sum_{(I,J) \in \mathcal{P}(S)} \mathbb{E}_{I_J}^\sigma \Delta_I^\sigma f \cdot \langle H(\mathbf{1}_S \sigma), \Delta_J^\omega \phi \rangle_\omega$$

$$(6.18) \quad A_3^4 \equiv \sum_{S \in \mathcal{S}} \sum_{(I,J) \in \mathcal{P}(S)} \mathbb{E}_{I_J}^\sigma \Delta_I^\sigma f \cdot \langle H(\mathbf{1}_{S \setminus I_J} \sigma), \Delta_J^\omega \phi \rangle_\omega$$

Recall that the three terms above are the neighbor, paraproduct, and stopping terms respectively.

The paraproduct term  $A_2^4$  is further decomposed, while we will prove in §9.5 and §9.1 respectively,

$$(6.19) \quad |A_1^4| \lesssim \mathcal{A}_2 \|f\|_{L^2(\sigma)} \|\phi\|_{L^2(\omega)},$$

$$(6.20) \quad |A_3^4| \lesssim \mathcal{F}_{\gamma, \varepsilon} \|f\|_{L^2(\sigma)} \|\phi\|_{L^2(\omega)}.$$

## 7. THE CARLESON MEASURE ESTIMATES

This section is devoted to the statement and proof of several Carleson measure estimates, designed with the considerations of the next section in mind. We collect them here, due to the common sets of techniques used to prove them.

The following technical Lemma encompasses many of the applications of the Energy Hypothesis and the stopping time definition. For an interval  $J \in \mathcal{D}^\omega$ , let

$$(7.1) \quad \widetilde{\mathbf{P}}_J^\omega \phi = \sum_{J' \in \mathcal{D}^\omega : J' \subset J} \langle \phi, h_{J'}^\omega \rangle_\sigma h_{J'}^\omega.$$

Note that this projection  $\tilde{\mathbf{P}}_J^\omega$  is onto the span of *all* Haar functions  $h_{J'}$  supported in the  $\mathcal{D}^\omega$ -interval  $J$ . By contrast,  $\mathbf{P}_S^\omega$  projects onto the span of those Haar functions  $h_J$  with  $J$  in the corona  $\mathcal{C}^\omega(S)$  where  $S$  is a stopping interval in the  $\mathcal{D}^\sigma$  grid.

**Lemma 7.2.** *Fix an interval  $I_0 \in \mathcal{D}^\sigma$  and let  $\hat{I}_0 \in \mathcal{S}$  be its  $\mathcal{S}$ -parent. Let  $\{I_r : r \geq 1\} \subset \mathcal{D}^\sigma$  be a strict subpartition of  $I_0$ . For  $r \geq 1$ , let  $\{J_{r,s} : s \geq 1\} \subset \mathcal{D}^\omega$  be a subpartition of  $I_r$  with  $|J_{r,s}| < 2^{-t}|I_r|$  for all  $r, s \geq 1$ , where  $t \geq r$  is the integer of Definition 4.2. We then have*

$$(7.3) \quad \sum_{r,s \geq 1} \left\| \tilde{\mathbf{P}}_{J_{r,s}}^\omega H(\mathbf{1}_{\hat{I}_0 \setminus I_r} \sigma) \right\|_{L^2(\omega)}^2 \lesssim 2^{-\gamma t} \mathcal{F}_{\gamma,\varepsilon}^2 \sigma(I_0).$$

*Proof.* We apply (6.7), and (1.16) to deduce the Lemma. We begin with (6.7) to obtain

$$\begin{aligned} \sum_{r,s \geq 1} \left\| \tilde{\mathbf{P}}_{J_{r,s}}^\omega H(\mathbf{1}_{\hat{I}_0 \setminus I_r} \sigma) \right\|_{L^2(\omega)}^2 &\lesssim \sum_{r,s \geq 1} \Phi(J_{r,s}, \hat{I}_0 \setminus I_r) \\ &\lesssim \sum_{r,s \geq 1} \Phi(J_{r,s}, \hat{I}_0 \setminus I_0) + \sum_{r,s \geq 1} \Phi(J_{r,s}, I_0 \setminus I_r), \end{aligned}$$

where the last inequality follows from the definition of  $\Phi$  and

$$\mathbf{P}(J, \mathbf{1}_{\hat{I}_0 \setminus I_r} \sigma) = \mathbf{P}(J, \mathbf{1}_{\hat{I}_0 \setminus I_0} \sigma) + \mathbf{P}(J, \mathbf{1}_{I_0 \setminus I_r} \sigma).$$

If  $I_0 \neq \hat{I}_0$ , we estimate the sum involving  $\hat{I}_0 \setminus I_0$  using the fact that  $\{J_{r,s}\}_{r,s \geq 1}$  is an  $\varepsilon$ -good subpartition of  $I_0$  (because the intervals  $J_{r,s}$  are good). We can thus use the third line in (1.14), and then the fact that (6.9) *fails* when  $I_0 \neq \hat{I}_0$ , to obtain

$$\sum_{r,s \geq 1} \Phi(J_{r,s}, \hat{I}_0 \setminus I_0) \lesssim \left\{ \sup_{s \geq 1} \left( \frac{|J_{r,s}|}{|I_0|} \right)^\gamma \right\} \Psi_{\gamma,\varepsilon}(I_0, \hat{I}_0) \lesssim 2^{-\gamma t} \mathcal{F}_{\gamma,\varepsilon}^2 \sigma(I_0).$$

Next, to estimate the sum involving  $I_0 \setminus I_r$ , we use the fact that  $\{J_{r,s}\}_{s \geq 1}$  is an  $\varepsilon$ -good subpartition of  $I_r$  for each  $r$  (again since the intervals  $J_{r,s}$  are good). We can thus use the third line in (1.14), and finally the Energy Hypothesis (1.16) to obtain

$$\sum_{r,s \geq 1} \Phi(J_{r,s}, I_0 \setminus I_r) \lesssim \sum_{r \geq 1} \left\{ \sup_{s \geq 1} \left( \frac{|J_{r,s}|}{|I_r|} \right)^\gamma \right\} \Psi_{\gamma,\varepsilon}(I_r, I_0) \lesssim 2^{-\gamma t} \mathcal{F}_{\gamma,\varepsilon}^2 \sigma(I_0).$$

This last estimate also proves the case  $I_0 = \hat{I}_0 \in \mathcal{S}$ . □

**Theorem 7.4.** *We have the following Carleson measure estimates for  $S \in \mathcal{S}$  and  $K \in \mathcal{D}^\sigma$ :*

$$(7.5) \quad \sum_{S' \in \mathcal{S}(S)} \sigma(S') \leq \frac{1}{4} \sigma(S) \quad \text{and} \quad \sum_{S \in \mathcal{S}: S \subsetneq K} \sigma(S) \leq \sigma(K),$$

$$(7.6) \quad \sum_{J \in \mathcal{C}^\omega(S): J \subset K, 2^r |J| < |K|} |\langle H(\mathbf{1}_S \sigma), h_J^\omega \rangle_\omega|^2 \lesssim (\mathcal{F}_{\gamma,\varepsilon}^2 + \mathcal{H}^2) \sigma(K).$$

**Remark 7.7.** *The Corona Decomposition and this last estimate can be compared to a general strategy for proving Carleson measure estimates. The Corona Decomposition is reminiscent of the sets in (3.11); the condition (7.5) can be compared to (3.10); and the condition (7.6) can be compared to (3.11).*

*Proof of (7.5).* Concerning the second inequality in (7.5), as is well known, it suffices to verify it for  $K = S_0 \in \mathcal{S}$ . And this case follows from the recursive application of the estimate first half of (7.5) to the interval  $S_0$  and all of its children in  $\mathcal{S}$ .

So we turn to the first half of (7.5). The intervals in the collection  $\mathcal{S}(S_0) = \{S_r : r \geq 1\}$  given in Definition 6.8 are pairwise disjoint and strictly contained in  $I_0$ . Each of them satisfies (6.9), so we can apply (1.16) to see that

$$(7.8) \quad \begin{aligned} \sum_{S \in \mathcal{S}(S_0)} \sigma(S) &= \sum_{r \geq 1} \sigma(S_r) \\ &\leq \frac{1}{4\mathcal{F}_{\gamma,\varepsilon}^2} \sum_{r \geq 1} \Psi_{\gamma,\varepsilon}(S_r, S_0) \leq \frac{1}{4} \sigma(S_0). \end{aligned}$$

□

*Proof of (7.6).* Fix  $S \in \mathcal{S}$  and  $K$ , which we can assume is a subset of  $S$ . If we apply the Hilbert transform to  $\sigma \mathbf{1}_K$ , as opposed to  $\sigma \mathbf{1}_S$ , we have by (3.4) for  $\omega$  and (1.3),

$$\sum_{\substack{J \in \mathcal{C}^\omega(S) \\ J \subset K, |J| < 2^{-r}|K|}} |\langle H(\mathbf{1}_K \sigma), h_J^\omega \rangle_\omega|^2 \leq \int_K |H(\mathbf{1}_K \sigma)|^2 \omega(dx) \leq \mathcal{H}^2 \sigma(K).$$

And so we consider the Hilbert transform applied to  $\sigma \mathbf{1}_{S \setminus K}$ , and show

$$(7.9) \quad \sum_{\substack{J \in \mathcal{C}^\omega(S) \\ J \subset K, |J| < 2^{-r}|K|}} |\langle H(\mathbf{1}_{S \setminus K} \sigma), h_J^\omega \rangle_\omega|^2 \lesssim \mathcal{F}_{\gamma,\varepsilon}^2 \sigma(K).$$

We can assume that  $K \subsetneq S$ , and that there is some  $J \in \mathcal{C}^\omega(S)$  with  $J \subset K$ . From this we see that  $K$  was *not* a stopping interval. That is, the interval  $K$  *must* fail (6.9).

Let  $\mathcal{J}$  denote the maximal intervals  $J \in \mathcal{C}^\omega(S)$  with  $J \subset K$  and  $|J| < 2^{-r}|K|$ . Using the notation of (7.1), we can use (7.3), with  $I' = K$ ,  $\hat{I} = S$ , and  $J \in \mathcal{J}$ . It gives us

$$\sum_{J \in \mathcal{J}} \|\tilde{P}_J^\omega H(\mathbf{1}_{S \setminus K} \sigma)\|_{L^2(\omega)}^2 \lesssim \sum_{J \in \mathcal{J}} \Phi(J, S \setminus K) \lesssim \mathcal{F}_{\gamma,\varepsilon}^2 \sigma(K).$$

The second inequality uses the fact that  $K$  fails (6.9). This proves (7.9). □

The following Carleson measure estimate, along with §9.1 and §8.5, are the three places where the Energy Hypothesis is used in this proof: It will provide the decay in the parameter  $t$  in (7.12). For all integers  $t \geq 0$ , we define for  $S \in \mathcal{S}$ , which are not maximal,

$$(7.10) \quad \alpha_t(S) \equiv \sum_{S': \pi_S^t(S')=S} \left\| \mathbf{P}_{S'}^\omega H(\sigma \mathbf{1}_{\pi_S^1(S) \setminus S}) \right\|_{L^2(\omega)}^2$$

Here, we are taking the projection  $H(\sigma \mathbf{1}_{\pi_S^1(S) \setminus S})$  associated to parts of the Corona decomposition which are ‘far below’  $S$ . We have this off-diagonal estimate.

**Theorem 7.11.** *The following Carleson measure estimate holds:*

$$(7.12) \quad \sum_{S \in \mathcal{S} : \pi_{\mathcal{D}^\sigma}^1(S) \subset K} \alpha_t(S) \lesssim 2^{-\gamma t} \mathcal{F}_{\gamma, \varepsilon}^2 \sigma(K), \quad K \in \mathcal{D}^\sigma.$$

*The implied constant is independent of the choice of interval  $K$  and  $t \geq 1$ .*

**Remark 7.13.** *In the estimate (7.12), we draw attention to the fact that the dyadic parent  $\pi_{\mathcal{D}^\sigma}^1(S)$  of  $S$  appears. Similar conditions will arise below, and it is essential to track them as the measures we are dealing with are not doubling. In fact, the role of the dyadic parents is revealed in the next proof: Use the negation of (6.9) when  $\pi_{\mathcal{D}^\sigma}^1(S) \notin \mathcal{S}$ , and otherwise use the Energy Condition.*

*Proof.* Our first task is to show that

$$\sum_{S \in \mathcal{S}(\hat{S})} \alpha_t(S) \leq 2^{-\gamma t} \mathcal{F}_{\gamma, \varepsilon}^2 \sigma(\hat{S}), \quad \hat{S} \in \mathcal{S}.$$

For the purposes of this proof, we will set  $\mathcal{S}_t(S) = \{S' \in \mathcal{S} : \pi_S^t(S') = S\}$ , using this notation for  $S \in \mathcal{S}(\hat{S})$ . We want to apply (6.7) to the expressions  $\alpha_t$ . To this end define

$$(7.14) \quad \mathcal{J}(S') \equiv \{J \in \mathcal{C}^\omega(S) : J \text{ is maximal w.r.t. } J \subset S', |J| < 2^{-r} |S'|\}.$$

It follows by definition that we have  $|J| < 2^{-r} |S'|$  for all  $J \in \mathcal{J}(S')$ . And, as all Haar functions have mean zero, we can apply (6.7). From this, we see that

$$\alpha_t(S) \lesssim \sum_{S' \in \mathcal{S}_t(S)} \sum_{J \in \mathcal{J}(S')} \Phi(J, \hat{S} \setminus S),$$

and so by the third line in (1.14),

$$\begin{aligned} \sum_{S \in \mathcal{S}(\hat{S})} \alpha_t(S) &\lesssim \sum_{S \in \mathcal{S}(\hat{S})} \sum_{S' \in \mathcal{S}_t(S)} \sum_{J \in \mathcal{J}(S')} \Phi(J, \hat{S} \setminus S) \\ &\lesssim 2^{-t\gamma} \sum_{S \in \mathcal{S}(\hat{S})} \Psi(S, \hat{S}) \lesssim 2^{-t\gamma} \mathcal{F}_{\gamma, \varepsilon}^2 \sigma(\hat{S}), \end{aligned}$$



where the final inequality follows from the assumed Energy Hypothesis (1.16).

Now fix  $K$  as in (7.12) and let  $\widehat{S} \in \mathcal{S}$  be the stopping interval such that  $K \in \mathcal{C}^\sigma(\widehat{S})$ . Let  $\mathcal{G}_1 \equiv \{S_i\}_i$  be the maximal intervals from  $\mathcal{S}$  that are strictly contained in  $K$ . Inductively define the  $(k+1)^{st}$  generation  $\mathcal{G}_{k+1}$  to consist of the maximal intervals from  $\mathcal{S}$  that are strictly contained in some  $k^{th}$  generation interval  $S \in \mathcal{G}_k$ . Inequality (7.15) shows that

$$\sum_{S \in \mathcal{G}_{k+1}} \alpha_t(S) \lesssim 2^{-t\gamma} \mathcal{F}_{\gamma, \varepsilon}^2 \sum_{S \in \mathcal{G}_k} \sigma(S).$$

We also have from (7.5) that

$$\sum_{k=1}^{\infty} \sum_{S \in \mathcal{G}_k} \sigma(S) \lesssim \sum_{S \in \mathcal{G}_1} \sigma(S) \leq \sigma(K).$$

This will be all we need in the case  $K = \widehat{S}$ , but when  $K \neq \widehat{S}$ , we will use Lemma 7.2 to control the first generation intervals  $S$  in  $\mathcal{G}_1$ :

$$\sum_{S \in \mathcal{G}_1} \alpha_t(S) \lesssim 2^{-t\eta} \sigma(K).$$

Indeed, we simply apply Lemma 7.2 with  $\widehat{I}_0 = \widehat{S}$ ,  $I_0 = K$ ,  $\{I_r\}_{r \geq 1} = \mathcal{G}_1$ , and  $\{J_{r,s}\}_{s \geq 1} = \bigcup_{S' \in \mathcal{S}_t(S)} \mathcal{J}(S')$ .

When  $K \neq \widehat{S}$  we finish with

$$\begin{aligned} \sum_{S \in \mathcal{S} : \pi_{\mathcal{D}\sigma}^1(S) \subset K} \alpha_t(S) &= \sum_{S \in \mathcal{G}_1} \alpha_t(S) + \sum_{k=1}^{\infty} \sum_{S \in \mathcal{G}_{k+1}} \alpha_t(S) \\ &\lesssim 2^{-t\eta} \sigma(K) + 2^{-t\eta} \mathcal{F}_{\gamma, \varepsilon}^2 \sum_{k=1}^{\infty} \sum_{S \in \mathcal{G}_k} \sigma(S) \\ &\lesssim 2^{-t\gamma} \mathcal{F}_{\gamma, \varepsilon}^2 \sigma(K), \end{aligned}$$

and when  $K = \widehat{S}$  we set  $\mathcal{G}_0 = \{\widehat{S}\}$  and estimate

$$\begin{aligned} \sum_{S \in \mathcal{S} : \pi_{\mathcal{D}\sigma}^1(S) \subset \widehat{S}} \alpha_t(S) &= \sum_{k=0}^{\infty} \sum_{S \in \mathcal{G}_{k+1}} \alpha_t(S) \lesssim 2^{-t\gamma} \mathcal{F}_{\gamma, \varepsilon}^2 \sum_{k=0}^{\infty} \sum_{S \in \mathcal{G}_k} \sigma(S) \\ &\lesssim 2^{-t\gamma} \mathcal{F}_{\gamma, \varepsilon}^2 \sigma(\widehat{S}). \end{aligned}$$

□

We need a Carleson measure estimate that is a common variant of (7.5) and (7.12). Define

$$(7.16) \quad \beta(S) \equiv \left\| \mathbf{P}_S^\omega H(\sigma \mathbf{1}_{\pi_{\mathcal{D}\sigma}^1(S)}) \right\|_{L^2(\omega)}^2.$$

**Theorem 7.17.** *We have the Carleson measure estimate*

$$(7.18) \quad \sum_{S \in \mathcal{S} : \pi_{\mathcal{D}\sigma}^1(S) \subset K} \beta(S) \lesssim (\mathcal{H}^2 + \mathcal{F}_{\gamma,\varepsilon}^2) \sigma(K)$$

*Proof.* Using the decomposition  $\pi_{\mathcal{D}\sigma}^1(S) = S \cup \{\pi_{\mathcal{D}\sigma}^1(S) \setminus S\}$ , we write  $\beta(S) \leq 2(\beta_1(S) + \beta_2(S))$  where

$$\begin{aligned} \beta_1(S) &\equiv \|\mathbf{P}_S^\omega H(\sigma \mathbf{1}_S)\|_{L^2(\omega)}^2, \\ \beta_2(S) &\equiv \left\| \mathbf{P}_S^\omega H(\sigma \mathbf{1}_{\pi_{\mathcal{D}\sigma}^1(S) \setminus S}) \right\|_{L^2(\omega)}^2. \end{aligned}$$

We certainly have  $\beta_1(S) \leq \mathcal{H}\sigma(S)$ , so that by (7.5), we need only consider the Carleson measure norm of the terms  $\beta_2(S)$ .

Fix an interval  $K$  of the form  $K = \pi_{\mathcal{D}}^1(S_0)$  for some  $S_0 \in \mathcal{S}$ . Let  $\mathcal{T}$  be the maximal intervals of the form  $\pi_{\mathcal{D}}^1(S) \subsetneq K$ , and for  $T \in \mathcal{T}$ , let  $\mathcal{S}(T)$  be all intervals  $S \in \mathcal{S}$  with  $S \subset T$  and  $S$  is maximal. Using the notation of (7.14) and (7.1), we can estimate

$$\begin{aligned} \sum_{T \in \mathcal{T}} \sum_{S \in \mathcal{S}(T)} \beta_2(S) &\lesssim \sum_{T \in \mathcal{T}} \sum_{S \in \mathcal{S}(T)} \sum_{J \in \mathcal{J}(S)} \left\| \tilde{\mathbf{P}}_J^\omega H(\sigma \mathbf{1}_{\pi_{\mathcal{D}\sigma}^1(S) \setminus S}) \right\|_{L^2(\omega)}^2 \\ &\lesssim \mathcal{F}_{\gamma,\varepsilon}^2 \sigma(K) \end{aligned}$$

Here, we have been careful to arrange the collections  $\mathcal{T}$ ,  $\mathcal{S}(T)$  and  $\mathcal{J}(S)$  so that (7.3) applies.

We argue that this inequality is enough to conclude the Lemma. Suppose that  $S' \in \mathcal{S}$ , with  $S' \subset K$ , but  $S'$  is not in any collection  $\mathcal{S}(T)$  for  $T \in \mathcal{T}$ . It follows that  $S' \subsetneq S$  for some  $S \in \mathcal{S}(T)$  and  $T \in \mathcal{T}$ . This implies that the Carleson measure estimate (7.5) will conclude the proof.  $\square$

A last Carleson measure estimate needed arises from the quantities

$$(7.19) \quad \gamma(S) \equiv \left\| \mathbf{P}_S^\omega H(\mathbf{1}_{\pi_S^1(S) \setminus \pi_{\mathcal{D}\sigma}^1(S)} \sigma) \right\|_{L^2(\omega)}^2$$

**Theorem 7.20.** *We have the estimate*

$$(7.21) \quad \sum_{S \in \mathcal{S} : \pi_{\mathcal{D}\sigma}^1(S) \subset K} \gamma(S) \lesssim \mathcal{F}_{\gamma,\varepsilon}^2 \sigma(K).$$

*Proof.* We can take  $K = \pi_{\mathcal{D}\sigma}^1(S_0)$  for some  $S_0 \in \mathcal{S}$ , and in addition, we can assume that  $K \notin \mathcal{S}$ , because otherwise we are applying the Hilbert transform to the zero function.

We repeat an argument from the previous proof. Details are omitted.  $\square$

## 8. THE PARAPRODUCTS

We continue to follow the line of argument in [Vol] and [NTV4] using similar notation for the benefit of the reader. The paraproduct term  $A_2^4$  is the central term in the proof. In this section, we reorganize the sum in (6.17) according to the Corona Decomposition: The essential point that must be accounted for is that for  $J \in \mathcal{C}^\omega(S)$  and  $J \subset I$ , we *need not* have  $I \in \mathcal{C}^\sigma(S)$ . On the other hand, it will be the case that  $I \in \mathcal{C}^\sigma(\pi_S^t(S))$  for some ancestor  $\pi_S^t(S)$  of  $S$ . The ancestor  $\pi_S^t(S)$  is only defined for  $1 \leq t \leq \rho(S)$ . (See Definition 6.8 for the definition of  $\rho(S)$ .) In fact, the sum splits into  $A_2^4 = A_1^5 + A_2^5$ , where

$$(8.1) \quad A_1^5 \equiv \sum_{S \in \mathcal{S}} \sum_{\substack{(I,J) \in \mathcal{P}(S) \\ J \in \mathcal{C}^\omega(S)}} \mathbb{E}_{I,J}^\sigma \Delta_I^\sigma f \cdot \langle H(\mathbf{1}_S \sigma), \Delta_J^\omega \phi \rangle_\omega ,$$

$$(8.2) \quad A_2^5 \equiv \sum_{S \in \mathcal{S} \setminus \{I^0\}} \sum_{t=1}^{\rho(S)} \sum_{\substack{(I,J) \in \mathcal{P}(\pi_S^t(S)) \\ J \in \mathcal{C}^\omega(S)}} \mathbb{E}_{I,J}^\sigma \Delta_I^\sigma f \cdot \left\langle H(\mathbf{1}_{\pi_S^t(S)} \sigma), \Delta_J^\omega \phi \right\rangle_\omega .$$

In  $A_1^5$ , we are treating the case where both  $I \in \mathcal{D}^\sigma$  and  $J$  are ‘controlled’ by the same stopping interval. ( $J$  is not ‘very far’ below  $I$ , as measured by the stopping intervals  $\mathcal{S}$ .) And, the point in the last line is that we are summing over  $J \in \mathcal{C}^\omega(S)$ , while the pair  $(I, J) \in \mathcal{P}(\pi_S^t(S))$ , where  $\pi_S^t(S)$  denotes the  $t$ -fold parent of  $S$  in the grid  $\mathcal{S}$ , see (3.1). This ancestor appears in two places, controlling the sum over  $I$ , and in the argument of the Hilbert transform.

We will prove

$$(8.3) \quad |A_1^5| \lesssim (\mathcal{H} + \mathcal{F}_{\gamma, \varepsilon}) \|f\|_{L^2(\sigma)} \|\phi\|_{L^2(\omega)} ,$$

while  $A_2^5$  will require further decomposition.

**8.1.  $A_1^5$ : The First Paraproduct.** We use the telescoping sum identities (3.5) and (3.6) to reorganize the sum in (8.1). Fix  $S \in \mathcal{S}$  and  $J \in \mathcal{C}^\omega(S)$ . The sum over  $I$  in (8.1) is (6.12).

$$(8.4) \quad \sum_{I : (I,J) \in \mathcal{P}(S)} \mathbb{E}_{I,J}^\sigma \Delta_I^\sigma f = \mathbb{E}_{I_{J,*}}^\sigma f - \mathbb{E}_{\pi_{\mathcal{D}}(S)}^\sigma f .$$

Here, we set  $I_{J,*}$  to be the minimal member of  $\mathcal{C}^\sigma(S)$  that contains  $J$ , and satisfies  $2^r |J| < |I|$ . Such an interval must exist as  $J$  is good. Thus, we can write

$$(8.5) \quad A_1^5 = \sum_{S \in \mathcal{S}} A_1^5(S) ,$$

$$(8.6) \quad A_1^5(S) \equiv \sum_{J \in \mathcal{C}^\omega(S)} \left( \mathbb{E}_{I_{J,*}}^\sigma f - \mathbb{E}_S^\sigma f \right) \cdot \langle H(\mathbf{1}_S \sigma), \Delta_J^\omega \phi \rangle_\omega .$$

The basic estimate here, and our first paraproduct style estimate is

**Proposition 8.7.** *We have the estimates*

$$(8.8) \quad |A_1^5(S)| \lesssim (\mathcal{H} + \mathcal{F}_{\gamma,\varepsilon}) \left\| \mathbf{P}_S^\sigma f - \mathbf{1}_S \mathbb{E}_{\pi_D(S)}^\sigma f \right\|_{L^2(\sigma)} \left\| \mathbf{P}_S^\omega \phi \right\|_{L^2(\omega)}, \quad S \in \mathcal{S}.$$

Here, the projections on the right are defined in (6.12).

*Proof.* We should reorganize the sum in a fashion consistent with paraproduct-type estimates. For  $I \in \mathcal{C}^\sigma(S)$ , let

$$\mathbf{Q}_I^\omega \phi \equiv \sum_{J \in \mathcal{C}^\omega(S): I_{J,*} = I} \Delta_J^\omega \phi.$$

Using the Cauchy-Schwartz inequality, and the fact that  $\mathbb{E}_I^\sigma f = \mathbb{E}_I^\sigma \mathbf{P}_S^\sigma f$  and  $\mathbf{Q}_I^\omega \phi = \mathbf{Q}_I^\omega \mathbf{P}_S^\omega \phi$  we see that

$$\begin{aligned} |A_1^5(S)| &= \left| \sum_{I \in \mathcal{C}^\sigma(S)} (\mathbb{E}_I^\sigma \mathbf{P}_S^\sigma f - \mathbb{E}_{\pi_D(S)}^\sigma f) \cdot \langle H(\mathbf{1}_S \sigma), \mathbf{Q}_I^\omega \mathbf{P}_S^\omega \phi \rangle_\omega \right| \\ &\leq \left[ \sum_{I \in \mathcal{C}^\sigma(S)} |\mathbb{E}_I^\sigma \mathbf{P}_S^\sigma f - \mathbb{E}_{\pi_D(S)}^\sigma f|^2 \cdot \|\mathbf{Q}_I^\omega H(\mathbf{1}_S \sigma)\|_{L^2(\omega)}^2 \sum_{I \in \mathcal{C}^\sigma(S)} \|\mathbf{Q}_I^\omega \mathbf{P}_S^\omega \phi\|_{L^2(\omega)}^2 \right]^{\frac{1}{2}} \\ &\leq \|\mathbf{P}_S^\omega \phi\|_{L^2(\omega)} \left[ \sum_{I \in \mathcal{C}^\sigma(S)} |\mathbb{E}_I^\sigma (\mathbf{P}_S^\sigma f - \mathbf{1}_S \mathbb{E}_{\pi_D(S)}^\sigma f)|^2 \cdot \|\mathbf{Q}_I^\omega H(\mathbf{1}_S \sigma)\|_{L^2(\omega)}^2 \right]^{1/2}. \end{aligned}$$

In view of the Carleson Embedding inequality, namely (3.8) and (3.9), this last factor is at most  $\left\| \mathbf{P}_S^\sigma f - \mathbf{1}_S \mathbb{E}_{\pi_D(S)}^\sigma f \right\|_{L^2(\sigma)}$  times the Carleson measure norm of the coefficients

$$\left\{ \|\mathbf{Q}_I^\omega H(\mathbf{1}_S \sigma)\|_{L^2(\omega)}^2 : I \in \mathcal{C}^\omega(S) \right\}.$$

But by the Plancherel formula (3.4) this is what is shown in (7.6) to be at most a constant multiple of  $\mathcal{H} + \mathcal{F}_{\gamma,\varepsilon}$ , so the proof is complete.  $\square$

To complete the estimate for  $A_1^5$ , from (8.5) and the observation that the projections on the right in (8.8) are essentially orthogonal, see (6.13), we can estimate

$$\begin{aligned} |A_1^5| &\lesssim (\mathcal{H} + \mathcal{F}_{\gamma,\varepsilon}) \sum_{S \in \mathcal{S}} \left\| \mathbf{P}_S^\sigma f - \mathbf{1}_S \mathbb{E}_{\pi_D(S)}^\sigma f \right\|_{L^2(\sigma)} \left\| \mathbf{P}_S^\omega \phi \right\|_{L^2(\omega)} \\ &\lesssim (\mathcal{H} + \mathcal{F}_{\gamma,\varepsilon}) \left( \sum_{S \in \mathcal{S}} \left\| \mathbf{P}_S^\sigma f - \mathbf{1}_S \mathbb{E}_{\pi_D(S)}^\sigma f \right\|_{L^2(\sigma)}^2 \sum_{S \in \mathcal{S}} \left\| \mathbf{P}_S^\omega \phi \right\|_{L^2(\omega)}^2 \right)^{1/2} \\ &\lesssim (\mathcal{H} + \mathcal{F}_{\gamma,\varepsilon}) \|f\|_{L^2(\sigma)} \|\phi\|_{L^2(\omega)}, \end{aligned}$$

since we have

$$\begin{aligned}
\sum_{S \in \mathcal{S}} \|\mathbf{1}_S \mathbb{E}_{\pi_{\mathcal{D}}(S)}^\sigma f\|_{L^2(\sigma)}^2 &= \sum_{S \in \mathcal{S}} \sigma(S) |\mathbb{E}_{\pi_{\mathcal{D}}(S)}^\sigma f|^2 \\
&\leq \sum_{S \in \mathcal{S}} \sigma(S) (\mathbb{E}_S^\sigma |f|)^2 \\
&\lesssim \|\mathcal{M}_\sigma f\|_{L^2(\sigma)}^2 \lesssim \|f\|_{L^2(\sigma)}^2.
\end{aligned}$$

Here, we should make an appeal to (7.5) in order to conclude that the maximal function  $\mathcal{M}_\sigma$  dominates the sum in (8.10). This is (8.3).

**8.2. The Remaining Paraproducts.** We repeat the analysis of (8.4), but for the term  $A_2^5$  defined in (8.2). Fix  $J$ , which must be a member of  $\mathcal{C}^\omega(S)$  for some  $S \in \mathcal{S} \setminus \{I^0\}$ . The sum over  $I$  in (8.2), as it turns out, is only a function of this  $S$ , and equals

$$\begin{aligned}
(8.12) \quad A_2^5(S) &\equiv \sum_{t=1}^{\rho(S)} \sum_{\substack{(I,J) \in \mathcal{P}(\pi_S^t(S)) \\ J \in \mathcal{C}^\omega(S)}} \mathbb{E}_{IJ}^\sigma \Delta_I^\sigma f \cdot \left\langle H(\mathbf{1}_{\pi_S^t(S)} \sigma), \Delta_J^\omega \phi \right\rangle_\omega \\
&= \sum_{t=1}^{\rho(S)} \sum_{J \in \mathcal{C}^\omega(S)} (\mathbb{E}_{\pi_{\mathcal{D}^\sigma}^2(\pi_S^{t-1}(S))}^\sigma f - \mathbb{E}_{\pi_{\mathcal{D}^\sigma}^1(\pi_S^t(S))}^\sigma f) \cdot \left\langle H(\mathbf{1}_{\pi_S^t(S)} \sigma), \Delta_J^\omega \phi \right\rangle_\omega.
\end{aligned}$$

We argue as follows. With  $J \in \mathcal{C}^\omega(S)$  fixed, the sum over  $I$  such that  $(I, J) \in \mathcal{P}(\pi_S^t(S))$  is only a function of  $S$  and  $t$ , and is a sum over consecutive intervals in the grid  $\mathcal{D}^\sigma$ . The smallest interval that contributes to the sum is  $\pi_{\mathcal{D}^\sigma}^2(\pi_S^{t-1}(S))$ , the *second* dyadic parent of  $\pi_S^{t-1}(S)$ , and the largest is  $\pi_{\mathcal{D}^\sigma}^1(\pi_S^t(S))$ . (Recall Definition 6.11. Also, these two intervals might be one and the same.)

In (8.12), the sum over  $J$  is independent of the sum over  $t$ . In the next steps, we concentrate on the sum over  $t$ . Below we add and subtract a cancellative term, to adjust for the second parent in (8.12).

$$\begin{aligned}
(8.13) \quad \tilde{A}_2^5(S) &= \sum_{t=1}^{\rho(S)} (\mathbb{E}_{\pi_{\mathcal{D}^\sigma}^2(\pi_S^{t-1}(S))}^\sigma f - \mathbb{E}_{\pi_{\mathcal{D}^\sigma}^1(\pi_S^{t-1}(S))}^\sigma f + \mathbb{E}_{\pi_{\mathcal{D}^\sigma}^1(\pi_S^{t-1}(S))}^\sigma f - \mathbb{E}_{\pi_{\mathcal{D}^\sigma}^1(\pi_S^t(S))}^\sigma f) \\
&\quad \times H(\mathbf{1}_{\pi_S^t(S)} \sigma) \\
&= \tilde{A}_{21}^5(S) + \tilde{A}_{22}^5(S),
\end{aligned}$$

$$(8.14) \quad \tilde{A}_{21}^5(S) \equiv \sum_{t=1}^{\rho(S)} \left( \mathbb{E}_{\pi_{\mathcal{D}^\sigma}^2(\pi_S^{t-1}(S))}^\sigma f - \mathbb{E}_{\pi_{\mathcal{D}^\sigma}^1(\pi_S^{t-1}(S))}^\sigma f \right) \cdot H(\mathbf{1}_{\pi_S^t(S)} \sigma),$$

$$(8.15) \quad \tilde{A}_{22}^5(S) \equiv \sum_{t=1}^{\rho(S)} \left( \mathbb{E}_{\pi_{\mathcal{D}^\sigma}^1(\pi_S^{t-1}(S))}^\sigma f - \mathbb{E}_{\pi_{\mathcal{D}^\sigma}^1(\pi_S^t(S))}^\sigma f \right) \cdot H(\mathbf{1}_{\pi_S^t(S)} \sigma).$$

The term  $\tilde{A}_{22}^5(S)$  in (8.15) is itself a telescoping sum, and so we can sum by parts to write

$$(8.16) \quad \tilde{A}_{22}^5(S) = \mathbb{E}_{\pi_{\mathcal{D}\sigma}^1(S)}^\sigma f \cdot H(\mathbf{1}_{\pi_S^1(S)}\sigma) + \sum_{t=1}^{\rho(S)} \mathbb{E}_{\pi_{\mathcal{D}\sigma}^1(\pi_S^t(S))}^\sigma f \cdot H(\mathbf{1}_{\pi_S^{t+1}(S) \setminus \pi_S^t(S)}\sigma).$$

Note that there is one term missing, but it has the expectation  $\mathbb{E}_{\pi_S^{\rho(S)}(S)}^\sigma f = \mathbb{E}_{I^0}^\sigma f$ , where  $I^0$  is the largest interval that we fixed at the beginning of the proof. In particular we have assumed that this expectation is zero.

We combine these steps, specifically the definition of  $A_2^5$  in (8.2) and the identities (8.13), (8.14), (8.15), and (8.16) to write  $A_2^5 = A_1^6 + A_2^6 + A_3^6$ , where

$$(8.17) \quad A_1^6(S) \equiv \sum_{t=1}^{\rho(S)} (\mathbb{E}_{\pi_{\mathcal{D}\sigma}^2(\pi_S^{t-1}(S))}^\sigma f - \mathbb{E}_{\pi_{\mathcal{D}\sigma}^1(\pi_S^{t-1}(S))}^\sigma f) \cdot \left\langle H(\mathbf{1}_{\pi_S^t(S)}\sigma), \mathbf{P}_S^\omega \phi \right\rangle_\omega,$$

$$(8.18) \quad A_2^6(S) \equiv \mathbb{E}_{\pi_{\mathcal{D}\sigma}^1(S)}^\sigma f \cdot \left\langle H(\mathbf{1}_{\pi_S^1(S)}\sigma), \mathbf{P}_S^\omega \phi \right\rangle_\omega,$$

$$A_3^6(S) \equiv \sum_{t=1}^{\rho(S)} \mathbb{E}_{\pi_{\mathcal{D}\sigma}^1(\pi_S^t(S))}^\sigma f \cdot \left\langle H(\mathbf{1}_{\pi_S^{t+1}(S) \setminus \pi_S^t(S)}\sigma), \mathbf{P}_S^\omega \phi \right\rangle_\omega.$$

Of these three expressions, the first  $A_1^6$  has cancellative terms on both  $f$  and  $\phi$ , hence it is not (yet) a paraproduct as such. The second  $A_2^6$  is a paraproduct, one that is very close in form to that of  $A_1^5$ , compare (8.6) and (8.18). The third term is a paraproduct, but looking at the support of the argument of the Hilbert transform, one sees that it is also degenerate, and we should obtain some additional decay in the parameter  $t$ , the ‘miraculous improvement of the Carleson property’ in Chapter 21.3 of [Vol] - see (8.21) below. We take up these estimates in the next subsections, passing from more intricate to less intricate.

In fact we will prove in §8.3, §8.4 and §8.5 respectively,

$$(8.19) \quad |A_3^6| \lesssim \mathcal{F}_{\gamma,\varepsilon} \|f\|_{L^2(\sigma)} \|\phi\|_{L^2(\omega)},$$

$$(8.20) \quad |A_2^6| \lesssim (\mathcal{H} + \mathcal{F}_{\gamma,\varepsilon}) \|f\|_{L^2(\sigma)} \|\phi\|_{L^2(\omega)},$$

$$|A_1^6| \lesssim (\mathcal{H} + \mathcal{F}_{\gamma,\varepsilon}) \|f\|_{L^2(\sigma)} \|\phi\|_{L^2(\omega)}.$$

In particular, the Energy Hypothesis enters into (8.19).

**8.3. The Term  $A_3^6$ .** Let us fix  $t$ , and define

$$A_3^6(S, t) \equiv \mathbb{E}_{\pi_{\mathcal{D}\sigma}^1(\pi_S^t(S))}^\sigma f \cdot \left\langle H(\mathbf{1}_{\pi_S^{t+1}(S) \setminus \pi_S^t(S)}\sigma), \mathbf{P}_S^\omega \phi \right\rangle_\omega, \quad S \in \mathcal{S}, \rho(S) \geq t.$$

$$A_3^6(t) \equiv \sum_{S \in \mathcal{S} : \rho(S) \geq t} A_3^6(S, t).$$

Here, we impose the restriction  $\rho(S) \geq t$  so that the  $t$ -fold parent of  $S$  is defined.

The estimate we prove is

$$(8.21) \quad |A_3^6(t)| \lesssim 2^{-\gamma t} \mathcal{F}_{\gamma, \varepsilon} \|f\|_{L^2(\sigma)} \|\phi\|_{L^2(\omega)}, \quad t \geq 1.$$

The constant  $\epsilon = \gamma/2 > 0$ . Clearly this proves (8.19) after summation on  $t \geq 1$ .

The projections  $\mathbf{P}_S^\omega$  are orthogonal, so we have

$$(8.22) \quad |A_3^6(t)| \leq \|\phi\|_{L^2(\omega)} \left[ \sum_{S \in \mathcal{S} : \rho(S) \geq t} \left| \mathbb{E}_{\pi_{\mathcal{D}^\sigma}^1}^\sigma(\pi_S^t(S)) f \right|^2 \left\| \mathbf{P}_S^\omega H(\mathbf{1}_{\pi_S^{t+1}(S) \setminus \pi_S^t(S)} \sigma) \right\|_{L^2(\omega)}^2 \right]^{1/2}.$$

Recalling the notation (7.10), the sum on the right in (8.22) is

$$\left[ \sum_{S \in \mathcal{S}} \alpha_t(S) \left| \mathbb{E}_{\pi_{\mathcal{D}^\sigma}^1}^\sigma(S) f \right|^2 \right]^{1/2}.$$

Therefore, to prove (8.21), we should verify that the Carleson measure norm of the coefficients  $\{\alpha_t(S) : S \in \mathcal{S}\}$  is at most  $C2^{-\gamma t} \mathcal{F}_{\gamma, \varepsilon}$ . But this is the content of Theorem 7.11, and so our proof is complete.

**8.4. The term  $A_2^6$ .** We certainly have  $\pi_{\mathcal{D}^\sigma}^1(S) \subset \pi_S^1(S)$ , so that it is natural to split term in (8.18) into two, namely writing  $\pi_S^1(S) = \pi_{\mathcal{D}^\sigma}^1(S) \cup \{\pi_S^1(S) \setminus \pi_{\mathcal{D}^\sigma}^1(S)\}$ , to give us

$$(8.23) \quad |A_1^7| \equiv \left| \sum_{S \in \mathcal{S}} \mathbb{E}_{\pi_{\mathcal{D}^\sigma}^1}^\sigma(S) f \cdot \left\langle H(\mathbf{1}_{\pi_{\mathcal{D}^\sigma}^1(S)} \sigma), \mathbf{P}_S^\omega \phi \right\rangle_\omega \right| \lesssim (\mathcal{H} + \mathcal{F}_{\gamma, \varepsilon}) \|f\|_{L^2(\sigma)} \|\phi\|_{L^2(\omega)},$$

$$(8.24) \quad |A_2^7| \equiv \left| \sum_{S \in \mathcal{S}} \mathbb{E}_{\pi_{\mathcal{D}^\sigma}^1}^\sigma(S) f \cdot \left\langle H(\mathbf{1}_{\pi_S^1(S) \setminus \pi_{\mathcal{D}^\sigma}^1(S)} \sigma), \mathbf{P}_S^\omega \phi \right\rangle_\omega \right| \lesssim \mathcal{F}_{\gamma, \varepsilon} \|f\|_{L^2(\sigma)} \|\phi\|_{L^2(\omega)}.$$

Together these prove (8.20). We treat them in turn.

Recalling the notation (7.16), we estimate

$$\begin{aligned} \left| \left\langle H(\mathbf{1}_{\pi_S^1(S)} \sigma), \mathbf{P}_S^\omega \phi \right\rangle_\omega \right| &= \left| \left\langle \mathbf{P}_S^\omega H(\mathbf{1}_{\pi_S^1(S)} \sigma), \mathbf{P}_S^\omega \phi \right\rangle_\omega \right| \\ &\leq \beta(S)^{1/2} \|\mathbf{P}_S^\omega \phi\|_{L^2(\omega)}. \end{aligned}$$

The latter projections are mutually orthogonal so we can estimate

$$\begin{aligned} |A_1^7| &\leq \left[ \sum_{S \in \mathcal{S}} \beta(S) \left| \mathbb{E}_{\pi_{\mathcal{D}^\sigma}^1}^\sigma(S) f \right|^2 \right]^{1/2} \|\phi\|_{L^2(\omega)} \\ &\lesssim (\mathcal{H} + \mathcal{F}_{\gamma, \varepsilon}) \|f\|_{L^2(\sigma)} \|\phi\|_{L^2(\omega)}. \end{aligned}$$

We have appealed to the Carleson measure estimate (7.18) to get the  $\|f\|_{L^2(\sigma)}$  term. This proves (8.23).

The argument for (8.24) is similar. Recalling the notation (7.19), we have

$$\begin{aligned} \left| \left\langle H(\mathbf{1}_{\pi_S^1(S) \setminus \pi_{\mathcal{D}\sigma}^1(S)} \sigma), \mathbf{P}_S^\omega \phi \right\rangle_\omega \right| &= \left| \left\langle \mathbf{P}_S^\omega H(\mathbf{1}_{\pi_S^1(S) \setminus \pi_{\mathcal{D}\sigma}^1(S)} \sigma), \mathbf{P}_S^\omega \phi \right\rangle_\omega \right| \\ &\leq \gamma(S)^{1/2} \|\mathbf{P}_S^\omega \phi\|_{L^2(\omega)} . \end{aligned}$$

We estimate

$$\begin{aligned} |A_2^7| &\leq \left[ \sum_{S \in \mathcal{S}} \gamma(S) \left| \mathbb{E}_{\pi_{\mathcal{D}\sigma}^1(S)}^\sigma f \right|^2 \right]^{1/2} \|\phi\|_{L^2(\omega)} \\ &\lesssim \mathcal{F}_{\gamma, \varepsilon} \|f\|_{L^2(\sigma)} \|\phi\|_{L^2(\omega)} . \end{aligned}$$

We have appealed to the Carleson measure estimate (7.21) to get the  $\|f\|_{L^2(\sigma)}$  term.

**8.5. The Term  $A_1^6$ .** In the definition of  $A_1^6(S)$ , see (8.17), note that the difference of expectations depends upon a single Haar coefficient, the one for the dyadic interval  $\pi_{\mathcal{D}\sigma}^2(\pi_S^t(S))$ . To be explicit, we will have the following equality

$$\mathbb{E}_{\pi_{\mathcal{D}\sigma}^2(S)}^\sigma f - \mathbb{E}_{\pi_{\mathcal{D}\sigma}^1(S)}^\sigma f = -\mathbb{E}_{\pi_{\mathcal{D}\sigma}^1(S)}^\sigma \Delta_{\pi_{\mathcal{D}\sigma}^2(S)}^\sigma, \quad S \in \mathcal{S} .$$

We reindex the sum defining  $A_1^6$  as follows. From (8.17), we write

$$\begin{aligned} -A_1^6(S) &= \sum_{t=1}^{\rho(S)} \mathbb{E}_{\pi_{\mathcal{D}\sigma}^1(\pi_S^{t-1}(S))}^\sigma \Delta_{\pi_{\mathcal{D}\sigma}^2(\pi_S^{t-1}(S))}^\sigma f \cdot \left\langle H(\mathbf{1}_{\pi_S^t(S)} \sigma), \mathbf{P}_S^\omega \phi \right\rangle_\omega = A_3^7(S) + A_4^7(S) , \\ (8.25) \quad A_3^7(S) &\equiv \sum_{t=1}^{\rho(S)} \mathbb{E}_{\pi_{\mathcal{D}\sigma}^1(\pi_S^{t-1}(S))}^\sigma \Delta_{\pi_{\mathcal{D}\sigma}^2(\pi_S^{t-1}(S))}^\sigma f \cdot \left\langle H(\mathbf{1}_{\pi_S^t(S) \setminus \pi_S^{t-1}(S)} \sigma), \mathbf{P}_S^\omega \phi \right\rangle_\omega \\ A_4^7(S) &\equiv \sum_{t=1}^{\rho(S)} \mathbb{E}_{\pi_{\mathcal{D}\sigma}^1(\pi_S^{t-1}(S))}^\sigma \Delta_{\pi_{\mathcal{D}\sigma}^2(\pi_S^{t-1}(S))}^\sigma f \cdot \left\langle H(\mathbf{1}_{\pi_S^{t-1}(S)} \sigma), \mathbf{P}_S^\omega \phi \right\rangle_\omega \end{aligned}$$

We argue that

$$(8.26) \quad \left| \sum_{S \in \mathcal{S} - \{S_0\}} A_3^7(S) \right| \lesssim \mathcal{F}_{\gamma, \varepsilon} \|f\|_{L^2(\sigma)} \|\phi\|_{L^2(\omega)} ,$$

$$(8.27) \quad \left| \sum_{S \in \mathcal{S} - \{S_0\}} A_4^7(S) \right| \lesssim \mathcal{F}_{\gamma, \varepsilon} \|f\|_{L^2(\sigma)} \|\phi\|_{L^2(\omega)} .$$

Indeed, the first inequality (8.26) is easier than the argument for (8.19), due to the extra orthogonality present with the Haar difference applied to  $f$  in (8.25). We omit the proof.



We turn to the proof of (8.27), and will need to appeal to our Energy Hypothesis again. Begin by reindexing the sum. We define

$$(8.28) \quad \begin{aligned} A_4^7(S, t) &\equiv \mathbb{E}_{\pi_{\mathcal{D}\sigma}^1(S)}^\sigma \Delta_{\pi_{\mathcal{D}\sigma}^2(S)}^\sigma f \cdot \sum_{\substack{S' \in \mathcal{S} \\ \pi^{t-1}(S')=S}} \langle H(\mathbf{1}_{S\sigma}), \mathbf{P}_{S'}^\omega \phi \rangle_\omega \\ \left| \sum_{S \in \mathcal{S}} A_4^7(S, t) \right| &\lesssim 2^{-\gamma t/2} \mathcal{F}_{\gamma, \varepsilon} \|f\|_{L^2(\sigma)} \|\phi\|_{L^2(\omega)}, \quad t \geq 1. \end{aligned}$$

(The decay in  $t$  is slightly worse in this case than in others.) Indeed, we first exploit the implicitly orthogonality in the sum. Note that we will have

$$\begin{aligned} \sum_{S \in \mathcal{S}} |\langle f, h_{\pi_{\mathcal{D}\sigma}^2(S)}^\sigma \rangle_\omega|^2 &\leq \|f\|_{L^2(\sigma)}^2, \\ \sum_{S \in \mathcal{S}} \left\| \sum_{S' \in \mathcal{S}: \pi^{t-1}(S')=S} \mathbf{P}_{S'}^\omega \phi \right\|_{L^2(\omega)}^2 &\leq \|\phi\|_{L^2(\omega)}^2. \end{aligned}$$

We also have from (3.2), that

$$\left| \mathbb{E}_{\pi_{\mathcal{D}\sigma}^1(S)}^\sigma h_{\pi_{\mathcal{D}\sigma}^2(S)}^\sigma \right| \leq |\pi_{\mathcal{D}\sigma}^1(S)|^{-1/2}$$

Combining these facts, we see that (8.28) follows from the estimate

$$(8.29) \quad \left\| \sum_{S' \in \mathcal{S}: \pi^{t-1}(S')=S} \mathbf{P}_{S'}^\omega H(\mathbf{1}_{\pi_S^{t-1}(S)} \sigma) \right\|_{L^2(\omega)}^2 \lesssim 2^{-\gamma t} (\mathcal{H}^2 + \mathcal{F}_{\gamma, \varepsilon}^2) |\pi_{\mathcal{D}\sigma}^1(S)|, \quad S \in \mathcal{S}, t \geq 1.$$

We turn to the proof of this last estimate. We will need geometric decay from two different sources. One is the geometric decay in (7.5), and the second is the application of the Energy Hypothesis, as in the proof of Theorem 7.11. Fix  $S \in \mathcal{S}$ , and integer  $u \simeq \frac{t-1}{2}$ , and let  $\mathcal{S}_u$  be those  $S' \in \mathcal{S}$  with  $\pi_S^u(S') = S$ . We have

$$\begin{aligned} \left\| \sum_{S' \in \mathcal{S}: \pi^{t-1}(S')=S} \mathbf{P}_{S'}^\omega H(\mathbf{1}_{\pi_S^{t-1}(S)} \sigma) \right\|_{L^2(\omega)}^2 &= \sum_{S' \in \mathcal{S}_u} B(S') \\ B(S') &\equiv \left\| \sum_{S'' \in \mathcal{S}: \pi^{t-1-u}(S'')=S'} \mathbf{P}_{S''}^\omega H(\mathbf{1}_{\pi_S^{t-1}(S)} \sigma) \right\|_{L^2(\omega)}^2 \end{aligned}$$

Now, in the definition of  $B(S')$ , we adjust the argument of the Hilbert transform, writing  $B(S') = B_1(S') + B_2(S')$ , where

$$B_1(S') \equiv \left\| \sum_{S'' \in \mathcal{S}: \pi^{t-1-u}(S'')=S'} P_{S''}^\omega H(\mathbf{1}_{\pi_S^{t-1}(S) \setminus S'} \sigma) \right\|_{L^2(\omega)}^2$$

$$B_2(S') \equiv \left\| \sum_{S'' \in \mathcal{S}: \pi^{t-1-u}(S'')=S'} P_{S''}^\omega H(\mathbf{1}_{S'} \sigma) \right\|_{L^2(\omega)}^2$$

Now, by the testing condition (1.3), we have

$$\begin{aligned} \sum_{S' \in \mathcal{S}_u} B_2(S') &\leq \mathcal{H}^2 \sum_{S' \in \mathcal{S}_u} \sigma(S') \\ &\leq 2^{-u/2} \mathcal{H}^2 \sigma(S) \leq 2^{-u/2} \mathcal{H}^2 \sigma(\pi_{\mathcal{D}^\sigma}^1(S)) \end{aligned}$$

where we have appealed to the Carleson measure property of the measure  $\sigma$  on the stopping cubes, more precisely (7.8), to deduce the last line. This proves half of (8.29).

We use the notation (7.14), and apply (7.3) to see that

$$\begin{aligned} \sum_{S' \in \mathcal{S}_u} B_1(S') &= \sum_{S' \in \mathcal{S}_u} \sum_{S'' \in \mathcal{S}: \pi^{t-1-u}(S'')=S'} \sum_{J \in \mathcal{J}(S'')} \|P_J^\omega H(\mathbf{1}_{\pi_S^{t-1}(S) \setminus S'} \sigma)\|_{L^2(\omega)}^2 \\ &\lesssim \mathcal{F}_{\gamma, \varepsilon}^2 2^{-\gamma t/2} \sigma(\pi_{\mathcal{D}^\sigma}^1(S)). \end{aligned}$$

This completes the proof of (8.29).

## 9. THE REMAINING ESTIMATES

We collect together the estimates claimed in earlier sections. The estimates in the first two subsections below are in [Vol], and the remaining three subsections essentially follow the arguments in [Vol] but using the Energy Hypothesis in §9.1.

**9.1.  $A_3^4$ : The Stopping Terms.** To control (6.18), and prove (6.20), it is important that we are dealing with the Energy Hypothesis (1.16).

We first claim that for  $S \in \mathcal{S}$  and  $s \geq 0$  an integer

$$\begin{aligned} (9.1) \quad A_3^4(S, s) &\equiv \sum_{(I, J) \in \mathcal{P}(S): |J|=2^{-s}|I|} |\mathbb{E}_{I, J}^\omega \Delta_I^\sigma f \cdot \langle H(\mathbf{1}_{S \setminus I, J} \sigma), \Delta_J^\omega \phi \rangle_\omega| \\ &\lesssim 2^{-\gamma s} \mathcal{F}_{\gamma, \varepsilon} F(S) \Lambda(S, s), \\ F(S)^2 &\equiv \sum_{I \in \mathcal{C}^\sigma(S)} |\langle f, h_I^\sigma \rangle_\sigma|^2 \\ \Lambda(S, s)^2 &\equiv \sum_{I \in \mathcal{C}^\sigma(S)} \sum_{J: (I, J) \in \mathcal{P}(S): |J|=2^{-s}|I|} |\langle \phi, h_J^\omega \rangle_\omega|^2. \end{aligned}$$

Indeed, apply Cauchy-Schwarz in the  $I$  variable above to obtain, and appeal to (3.2) to see that

$$A_3^4(S, s) \leq F(S) \left[ \sum_{I \in \mathcal{C}^\sigma(S)} \left( \sum_{J : (I, J) \in \mathcal{P}(S) : |J| = 2^{-s}|I|} \frac{1}{\sigma(I_J)^{1/2}} |\langle H(\mathbf{1}_{S \setminus I_J} \sigma), \Delta_J^\omega \phi \rangle_\omega| \right)^2 \right]^{\frac{1}{2}},$$

We can then estimate the sum inside the braces by

$$\begin{aligned} & \sum_{I \in \mathcal{C}^\sigma(S)} \sum_{J : (I, J) \in \mathcal{P}(S) : |J| = 2^{-s}|I|} |\langle \phi, h_J^\omega \rangle_\omega|^2 \\ & \times \sum_{J : (I, J) \in \mathcal{P}(S) : |J| = 2^{-s}|I|} \frac{1}{\sigma(I_J)} \cdot |\langle H(\mathbf{1}_{S \setminus I_J} \sigma), h_J^\omega \phi \rangle_\omega|^2 \\ & \lesssim \Lambda(S, s)^2 \cdot A(S, s) \\ A(S, s) & \equiv \sup_{I \in \mathcal{C}^\sigma(S)} \sum_{J : (I, J) \in \mathcal{P}(S) : |J| = 2^{-s}|I|} \sigma(I_J)^{-1} \cdot |\langle H(\mathbf{1}_{S \setminus I_J} \sigma), h_J^\omega \phi \rangle_\omega|^2. \end{aligned}$$

We turn to the analysis of the supremum in last display. We denote the two children of  $I$  by  $I_\theta$  for  $\theta \in \{-, +\}$ . Using (6.6) and then the third inequality in (1.14), we have

$$\begin{aligned} A(S, s) & \lesssim \sup_{I \in \mathcal{C}^\sigma(S)} \sup_{\theta \in \{-, +\}} \sigma(I_\theta)^{-1} \sum_{J : (I, J) \in \mathcal{P}(S) : I_J = I_\theta, |J| = 2^{-s}|I|} \Phi(J, S \setminus I_\theta) \\ & \lesssim \sup_{I \in \mathcal{C}^\sigma(S)} \sup_{\theta \in \{-, +\}} \sigma(I_\theta)^{-1} 2^{-\gamma s} \Psi_{\gamma, \varepsilon}(I_\theta, S) \\ & \lesssim 2^{-\gamma s} \mathcal{F}_{\gamma, \varepsilon}^2. \end{aligned}$$

The third inequality is the one for which the definition of stopping intervals was designed to deliver: From Definition 6.11, as  $(I, J) \in \mathcal{P}(S)$ , we have that  $S$  is the  $\mathcal{S}$ -parent of  $I_J$ , hence  $I_J$  was *not* a stopping interval, that is (6.9) does not hold, delivering the estimate above.

We clearly have from (3.4) that

$$\sum_{S \in \mathcal{S}} F(S)^2 \leq \sum_I |\langle f, h_I^\sigma \rangle_\sigma|^2 = \|f\|_{L^2(\sigma)}^2.$$

And so we have from (9.1),

$$\begin{aligned} \sum_{S \in \mathcal{S}} \sum_{s=0}^{\infty} A_3^4(S, s) & \lesssim \mathcal{F}_{\gamma, \varepsilon} \|f\|_{L^2(\sigma)} \left( \sum_{S \in \mathcal{S}} \sum_{s=0}^{\infty} 2^{-\gamma s} \Lambda(S, s)^2 \right)^{1/2} \\ & \lesssim \mathcal{F}_{\gamma, \varepsilon} \|f\|_{L^2(\sigma)} \|\phi\|_{L^2(\omega)}. \end{aligned}$$

**9.2.  $A_1^2$ : Diagonal Short Range Terms.** To prove (5.3), let us recall the definition (5.2). The pairs of intervals  $I, J$  arise from the dyadic grids  $\mathcal{D}^\sigma$  and  $\mathcal{D}^\omega$  respectively. But these grids share a common set of endpoints of the intervals. And the intervals  $I, J$  have comparable lengths,  $2^{-r}|I| \leq |J| \leq |I|$ . Accordingly, these pairs of intervals satisfy the conditions of the weak boundedness condition (2.7). A Haar function  $h_I^\sigma$  is a linear combination of its children, and the children of  $I$  and  $J$  also satisfy the weak boundedness condition (2.7). From this, we see that

$$|\langle H(\sigma h_I^\sigma), h_J^\omega \rangle_\omega| \leq 4\mathcal{W}, \quad (I, J) \in \mathcal{A}_1^2.$$

The Schur test easily implies that

$$A_1^2 \leq 4\mathcal{W} \sum_{I \in \mathcal{D}^\sigma} |\langle f, h_I^\sigma \rangle_\sigma| \sum_{J : (I, J) \in \mathcal{A}_1^2} |\langle \phi, h_J^\omega \rangle_\omega| \lesssim \mathcal{W} \|f\|_{L^2(\sigma)} \|\phi\|_{L^2(\omega)}.$$

**9.3.  $A_2^2$ : The Long Range Term.** We prove the estimate (5.4). Recall that the pairs of intervals  $I, J$  in question satisfy  $|J| \leq |I|$  and  $\text{dist}(I, J) \geq |I|$ . The hypotheses of (6.6) are in force, in particular (6.5) holds.

We observe that the Energy Lemma can be applied to estimate the inner product  $\langle H(h_I^\sigma \sigma), h_J^\omega \rangle_\omega$ . To see this, note that  $h_I^\sigma$  is constant on each child  $I_\pm$ . So, take a child  $I_\theta$ , and apply the Energy Lemma with the largest interval  $\hat{I}$  taken to be

$$\hat{I} = \text{hull} \left[ I_\theta, \left( \frac{|I|}{|J|} \right)^{1-\varepsilon} J \right].$$

Here  $\lambda J$  means the interval with the same center as  $J$  and length equal to  $\lambda|J|$ . The two intervals  $I_\theta$  and  $\left( \frac{|I|}{|J|} \right)^{1-\varepsilon} J$  are disjoint. We take  $I' \subset \hat{I}$  so that  $\hat{I} \setminus I' = I_\theta$ . Then, the Energy Lemma (6.6) and (3.2) apply to give us the estimate below.

$$\begin{aligned} (9.2) \quad \beta(I, J) &\equiv \left| \sum_{\theta} \langle H(\mathbf{1}_{I_\theta} h_I^\sigma \sigma), h_J^\omega \rangle_\omega \right| \leq |\mathbb{E}_{I_\theta}^\sigma h_I^\sigma| \sum_{\theta} |\langle H(\mathbf{1}_{I_\theta} \sigma), h_J^\omega \rangle_\omega| \\ &\lesssim \sum_{\theta} \left[ \frac{\omega(J)}{\sigma(I_\theta)} \right]^{1/2} \mathbf{E}(J, \omega) \mathbf{P}(J, \mathbf{1}_{\hat{I} \setminus I'} \sigma) \\ &\lesssim \sum_{\theta} \omega(J)^{1/2} \sigma(I_\theta)^{1/2} \cdot \frac{|J|}{\text{dist}(I, J)^2}. \end{aligned}$$

We have used the trivial inequalities  $\mathbf{E}(\omega, J) \leq 1$  and  $\mathbf{P}(J, \mathbf{1}_{I_\theta} \sigma) \leq \frac{|J|}{\text{dist}(I, J)^2} \sigma(I_\theta)$ .

We may assume that  $\|f\|_{L^2(\sigma)}^2 = \|\phi\|_{L^2(\omega)}^2 = 1$ . We then estimate

$$|A_2^2| \leq \sum_I \sum_{J : |J| \leq |I| : \text{dist}(I, J) \geq |I|} |\langle f, h_I^\sigma \rangle_\sigma| |\beta(I, J)| |\langle \phi, h_J^\omega \rangle_\omega|$$

$$\begin{aligned}
&\lesssim \sum_I \sum_{J : |J| \leq |I| : \text{dist}(I, J) \geq |I|} |\langle f, h_I^\sigma \rangle_\sigma| \sigma(I)^{\frac{1}{2}} \frac{|J|}{\text{dist}(I, J)^2} \omega(J)^{\frac{1}{2}} |\langle \phi, h_J^\omega \rangle_\omega| \\
&\lesssim \sum_I |\langle f, h_I^\sigma \rangle_\sigma|^2 \sum_{J : |J| \leq |I| : \text{dist}(I, J) \geq |I|} \left( \frac{|J|}{|I|} \right)^{-\delta} \sigma(I)^{\frac{1}{2}} \frac{|J|}{\text{dist}(I, J)^2} \omega(J)^{\frac{1}{2}} \\
&\quad + \sum_J |\langle \phi, h_J^\omega \rangle_\omega|^2 \sum_{I : |J| \leq |I| : \text{dist}(I, J) \geq |I|} \left( \frac{|J|}{|I|} \right)^{\delta} \sigma(I)^{\frac{1}{2}} \frac{|J|}{\text{dist}(I, J)^2} \omega(J)^{\frac{1}{2}},
\end{aligned}$$

where we have inserted the gain and loss factors  $\left( \frac{|J|}{|I|} \right)^{\pm\delta}$  with  $0 < \delta < 1$  to facilitate application of Schur's test. For each fixed  $I$  we have

$$\begin{aligned}
&\sum_{J : |J| \leq |I| : \text{dist}(I, J) \geq |I|} \left( \frac{|J|}{|I|} \right)^{\delta} \sigma(I)^{\frac{1}{2}} \frac{|J|}{\text{dist}(I, J)^2} \omega(J)^{\frac{1}{2}} \\
&\lesssim \sigma(I)^{\frac{1}{2}} \sum_{k=0}^{\infty} 2^{-k\delta} \left( \sum_{J : 2^k |J| = |I| : \text{dist}(I, J) \geq |I|} \frac{|J|}{\text{dist}(I, J)^2} \omega(J) \right)^{\frac{1}{2}} \\
&\quad \times \left( \sum_{J : 2^k |J| = |I| : \text{dist}(I, J) \geq |I|} \frac{|J|}{\text{dist}(I, J)^2} \right)^{\frac{1}{2}},
\end{aligned}$$

which is bounded by

$$\sum_{k=0}^{\infty} 2^{-k\delta} \left( \frac{\sigma(I)}{|I|} \mathbf{P}(I, \omega) \right)^{\frac{1}{2}} \lesssim \mathcal{A}_2,$$

if  $\delta > 0$ . For each fixed  $J$  we have

$$\begin{aligned}
&\sum_{I : |J| \leq |I| : \text{dist}(I, J) \geq |I|} \left( \frac{|J|}{|I|} \right)^{-\delta} \sigma(I)^{\frac{1}{2}} \frac{|J|}{\text{dist}(I, J)^2} \omega(J)^{\frac{1}{2}} \\
&\lesssim \omega(J)^{\frac{1}{2}} \sum_{k=0}^{\infty} 2^{-k(1-\delta)} \sum_{I : 2^k |J| = |I| : \text{dist}(I, J) \geq |I|} \frac{|I|}{\text{dist}(I, J)^2} \sigma(I)^{\frac{1}{2}} \\
&\lesssim \omega(J)^{\frac{1}{2}} \sum_{k=0}^{\infty} 2^{-k(1-\delta)} \left( \sum_{I : 2^k |J| = |I| : \text{dist}(I, J) \geq |I|} \frac{|I|}{\text{dist}(I, J)^2} \sigma(I) \right)^{\frac{1}{2}} \\
&\quad \times \left( \sum_{I : 2^k |J| = |I| : \text{dist}(I, J) \geq |I|} \frac{|I|}{\text{dist}(I, J)^2} \right)^{\frac{1}{2}},
\end{aligned}$$

which is bounded by

$$\begin{aligned} & \omega(J)^{\frac{1}{2}} \sum_{k=0}^{\infty} 2^{-k(1-\delta)} \mathbf{P}(2^k J, \sigma)^{\frac{1}{2}} \left( \frac{1}{|2^k J|} \right)^{\frac{1}{2}} \\ & \lesssim \sum_{k=0}^{\infty} 2^{-k(1-\delta)} \left( \frac{\omega(2^k J)}{|2^k J|} \mathbf{P}(2^k J, \sigma) \right)^{\frac{1}{2}} \lesssim \mathcal{A}_2, \end{aligned}$$

if  $\delta < 1$ . With any fixed  $0 < \delta < 1$  we obtain from the inequalities above that

$$\begin{aligned} |A_2^2| & \lesssim \sum_I |\langle f, h_I^\sigma \rangle_\sigma|^2 \mathcal{A}_2 + \sum_J |\langle \phi, h_J^\omega \rangle_\omega|^2 \mathcal{A}_2 \\ & = \left( \|f\|_{L^2(\sigma)}^2 + \|\phi\|_{L^2(\omega)}^2 \right) \mathcal{A}_2 = 2\mathcal{A}_2 \|f\|_{L^2(\sigma)}^2 \|\phi\|_{L^2(\omega)}^2, \end{aligned}$$

since we assumed  $\|f\|_{L^2(\sigma)}^2 = \|\phi\|_{L^2(\omega)}^2 = 1$ .

9.4.  **$A_1^3$  The Mid-Range Term.** We control the term associated with (5.5), namely we prove (5.7). For integers  $s \geq r$ , set

$$\begin{aligned} A_1^3(s) & \equiv \sum_I \sum_{J : 2^s |J|=|I| : \text{dist}(I, J) \leq |I|, I \cap J = \emptyset} |\langle f, h_I^\sigma \rangle_\sigma \langle H(h_I^\sigma \sigma), h_J^\omega \rangle_\omega \langle \phi, h_J^\omega \rangle_\omega| \\ & \lesssim \|f\|_{L^2(\sigma)} \left[ \sum_I \left( \sum_{J : 2^s |J|=|I| : \text{dist}(I, J) \leq |I|, I \cap J = \emptyset} |\langle H(h_I^\sigma \sigma), h_J^\omega \rangle_\omega \langle \phi, h_J^\omega \rangle_\omega| \right)^2 \right]^{1/2} \\ & \lesssim \Lambda(s) \|f\|_{L^2(\sigma)} \|\phi\|_{L^2(\omega)}, \\ \Lambda(s)^2 & \equiv 2^s \sup_I \sum_{J : 2^s |J|=|I| : \text{dist}(I, J) \leq |I|, I \cap J = \emptyset} |\langle H(h_I^\sigma \sigma), h_J^\omega \rangle_\omega|^2 \end{aligned}$$

since, by (3.4),

$$\sum_I \sum_{J : 2^s |J|=|I| : \text{dist}(I, J) \leq |I|, I \cap J = \emptyset} |\langle \phi, h_J^\omega \rangle_\omega|^2 = 2^s \sum_I \sum_J |\langle \phi, h_J^\omega \rangle_\omega|^2 = 2^s \|\phi\|_{L^2(\omega)}^2.$$

Due to the ‘local’ nature of the sum in  $J$ , we have thus gained a small improvement in the Schur test to derive the last line.

But  $J$  is good, so that (6.6) applies to each child  $I_\pm$  of  $I$  as in (9.2) above. Hence, we have using (2.18) that

$$\begin{aligned} \Lambda(s)^2 & \lesssim \sup_I 2^s \sum_{\theta} \sum_{J : 2^s |J|=|I| : \text{dist}(I, J) \leq |I|, I \cap J = \emptyset} \frac{\omega(J)}{\sigma(I_\theta)} \cdot \mathbf{E}(J, \omega)^2 \cdot \mathbf{P}(J, \mathbf{1}_{I_\theta} \sigma)^2 \\ & \lesssim \sup_I 2^s \sum_{\theta} \sum_{J : 2^s |J|=|I| : \text{dist}(I, J) \leq |I|, I \cap J = \emptyset} \frac{\omega(J)}{\sigma(I_\theta)} \left( \frac{|J|}{|I|} \right)^{2-2\varepsilon} \cdot \mathbf{P}(I_\theta, \mathbf{1}_{I_\theta} \sigma)^2 \end{aligned}$$

$$\begin{aligned}
&\lesssim \sup_I 2^s 2^{-s(2-2\varepsilon)} \sum_{\theta} \frac{\sigma(I_{\theta})}{|I|^2} \sum_{J: 2^s|J|=|I|: \text{dist}(I,J) \leq |I|, I \cap J = \emptyset} \omega(J) \\
&\lesssim 2^{-(1-2\varepsilon)s} \mathcal{A}_2.
\end{aligned}$$

This is clearly a summable estimate in  $s \geq r$ , so the proof of (5.7) is complete.

**9.5.  $A_1^4$ : The Neighbor Terms.** The neighbor terms are defined in (5.6), (6.1), (6.16), and we are to prove (6.19). To recall,  $I \in \mathcal{D}^{\sigma}$ ,  $J \in \mathcal{D}^{\omega}$  is contained in  $I$ , with  $|J| < 2^{-r}|I|$ , and  $I_J$  is the child of  $I$  that contains  $J$ .

Fix  $\theta \in \{-, +\}$ , and an integer  $s \geq r$ . Below we will use the convention that  $I \setminus I_{\theta} = I_{-\theta}$ . The inner product to be estimated is that in (6.16):

$$\begin{aligned}
\langle H(\mathbf{1}_{I_{-\theta}} \sigma \Delta_I^{\sigma} f), \Delta_J^{\omega} \phi \rangle_{\omega} &= \langle (\mathbf{1}_{I_{-\theta}} \sigma \Delta_I^{\sigma} f, H(\omega \Delta_J^{\omega} \phi))_{\sigma} \\
&= \mathbb{E}_{I_{-\theta}}^{\sigma} \Delta_I^{\sigma} f \cdot \langle (\mathbf{1}_{I_{-\theta}} \sigma, H(\omega \Delta_J^{\omega} \phi))_{\sigma} \\
&= \mathbb{E}_{I_{-\theta}}^{\sigma} \Delta_I^{\sigma} f \cdot \langle H(\mathbf{1}_{I_{-\theta}} \sigma), \Delta_J^{\omega} \phi \rangle_{\sigma}.
\end{aligned}$$

Use  $\|\Delta_J^{\omega} \phi\|_{L^2(\omega)} = |\langle \phi, h_J^{\omega} \rangle_{\omega}|$  and  $\frac{|J|}{|I_{\theta}|} = 2^{-s}$  in the Energy Lemma with  $J \subset I_{\theta} \subset I$  to obtain

$$\begin{aligned}
|\langle H(\mathbf{1}_{I_{-\theta}} \sigma), \Delta_J^{\omega} \phi \rangle_{\omega}| &\lesssim |\langle \phi, h_J^{\omega} \rangle_{\omega}| \omega(J)^{\frac{1}{2}} \cdot E(J, \omega) \cdot \mathbf{P}(J, \mathbf{1}_{I_{-\theta}} \sigma) \\
&\lesssim |\langle \phi, h_J^{\omega} \rangle_{\omega}| \omega(J)^{\frac{1}{2}} \cdot 2^{-(1-\varepsilon)s} \mathbf{P}(I_{-\theta}, \mathbf{1}_{I_{-\theta}} \sigma)
\end{aligned}$$

Here, we are using  $E(J, \omega) \leq 1$  and (2.18), which inequality applies since  $J \subset I \setminus I_{-\theta}$ .

In the sum below, we keep the length of the intervals  $J$  fixed, and assume that  $J \subset I_{\theta}$ . We estimate

$$\begin{aligned}
A_1^4(I, \theta, s) &\equiv \sum_{J: 2^s|J|=|I|: J \subset I_{\theta}} |\langle H(\mathbf{1}_{I_{-\theta}} \sigma \Delta_I^{\sigma} f), \Delta_J^{\omega} \phi \rangle_{\omega}| \\
&\leq 2^{-(1-\varepsilon)s} |\mathbb{E}_{I_{-\theta}}^{\sigma} \Delta_I^{\sigma} f| \mathbf{P}(I_{\theta}, \mathbf{1}_{I_{-\theta}} \sigma) \sum_{J: 2^s|J|=|I|: J \subset I_{\theta}} |\langle \phi, h_J^{\omega} \rangle_{\omega}| \omega(J)^{1/2} \\
&\leq 2^{-(1-\varepsilon)s} |\mathbb{E}_{I_{-\theta}}^{\sigma} \Delta_I^{\sigma} f| \mathbf{P}(I_{\theta}, \mathbf{1}_{I_{-\theta}} \sigma) \omega(I_{\theta})^{1/2} \Lambda(I, \theta, s), \\
\Lambda(I, \theta, s)^2 &\equiv \sum_{J: 2^s|J|=|I|: J \subset I_{\theta}} |\langle \phi, h_J^{\omega} \rangle_{\omega}|^2.
\end{aligned}$$

The last line follows upon using the Cauchy-Schwartz inequality.

Using (3.2), we have

$$|\mathbb{E}_{I_{-\theta}}^{\sigma} \Delta_I^{\sigma} f| \leq |\langle f, h_I^{\sigma} \rangle_{\sigma}| \cdot \sigma(I_{-\theta})^{-1/2}$$

And so, we can estimate  $A_1^4(I, \theta, s)$  as follows, in which we use the  $A_2$  hypothesis (1.6):

$$\begin{aligned}
A_1^4(I, \theta, s) &\lesssim 2^{-(1-\varepsilon)s} |\langle f, h_I^{\sigma} \rangle_{\sigma}| \Lambda(I, \theta, s) \cdot \sigma(I_{-\theta})^{-1/2} \mathbf{P}(I_{\theta}, \mathbf{1}_{I_{-\theta}} \sigma) \omega(I_{\theta})^{1/2} \\
&\lesssim \mathcal{A}_2 2^{-(1-\varepsilon)s} |\langle f, h_I^{\sigma} \rangle_{\sigma}| \Lambda(I, \theta, s),
\end{aligned}$$

since  $\mathbf{P}(I_\theta, \mathbf{1}_{I_\theta} \sigma) \lesssim \frac{\sigma(I_\theta)}{|I_\theta|}$  shows that

$$\sigma(I_\theta)^{-1/2} \mathbf{P}(I_\theta, \mathbf{1}_{I_\theta} \sigma) \omega(I_\theta)^{1/2} \lesssim \frac{\sigma(I_\theta)^{1/2} \omega(I_\theta)^{1/2}}{|I_\theta|} \lesssim \mathcal{A}_2,$$

A straight forward application of Cauchy-Schwartz then shows that

$$\sum_I A_1^4(I, \theta, s) \lesssim \mathcal{A}_2 2^{-(1-\varepsilon)s} \|f\|_{L^2(\sigma)} \|\Lambda(I, \theta, s)\|_{L^2(\omega)}.$$

This estimate is summable in  $\theta \in \{-, +\}$  and  $s \geq r$ , so the proof of (6.19) is complete.

## 10. A COUNTEREXAMPLE TO THE PIVOTAL CONDITIONS

We exhibit a weight pair  $(\omega, \sigma)$  that illustrates the nature of the Energy Condition, and the subtlety of the two weight problem in general. In particular it shows that the Pivotal Conditions are not necessary for the two weight inequality (1.2).

**Theorem 10.1.** *There is a weight pair  $(\omega, \sigma)$  which satisfies the two weight inequality (1.2), and fails the dual Pivotal Condition, namely (1.7) with the roles of  $\omega$  and  $\sigma$  reversed.*

Thus, this pair of weights satisfy the two weight inequality, but would not be included in the analysis of [NTV4]. We prove this result by appealing to our Theorem 1.17. In the course of the construction, we will see that one can make seemingly small modifications of the example measure  $\sigma$ , and in so doing violate the  $L^2$  inequality.

The plan of the proof of the Theorem is to (1) construct the pair of weights, and then to verify (2) the assertions on the Hybrid Conditions, (3) the  $A_2$  condition and (4) the two testing conditions (1.3) and (1.4). We take up these steps in the subsections below.

**10.1. Construction of the Pair of Weights.** Recall the middle-third Cantor set  $\mathbf{E}$  and Cantor measure  $\omega$  on the closed unit interval  $I_1^0 = [0, 1]$ . At the  $k^{\text{th}}$  generation in the construction, there is a collection  $\{I_j^k\}_{j=1}^{2^k}$  of  $2^k$  pairwise disjoint closed intervals of length  $|I_j^k| = \frac{1}{3^k}$ . With  $K_k = \bigcup_{j=1}^{2^k} I_j^k$ , the Cantor set is defined by  $\mathbf{E} = \bigcap_{k=1}^{\infty} K_k = \bigcap_{k=1}^{\infty} \left( \bigcup_{j=1}^{2^k} I_j^k \right)$ . The Cantor measure  $\omega$  is the unique probability measure supported in  $\mathbf{E}$  with the property that it is equidistributed among the intervals  $\{I_j^k\}_{j=1}^{2^k}$  at each scale  $k$ , i.e.

$$\omega(I_j^k) = 2^{-k}, \quad k \geq 0, 1 \leq j \leq 2^k.$$

We will define three measures  $\sigma, \dot{\sigma}, \ddot{\sigma}$ . We denote the removed open middle third of  $I_j^k$  by  $G_j^k$ . The three measures, restricted to an interval  $G_j^k$  will be a point mass with weight that is only a function of  $k$ . The only distinction will be the location of the point mass.



Let  $z_j^k \in G_j^k$  be the center of the interval  $G_j^k$ , which is also the center of the interval  $I_j^k$ . Now we define

$$(10.2) \quad \dot{\sigma} = \sum_{k,j} s_j^k \delta_{z_j^k},$$

where the sequence of positive numbers  $s_j^k$  is chosen to satisfy the following precursor of the  $A_2$  condition:

$$\frac{s_j^k \omega(I_j^k)}{|I_j^k|^2} = 1, \quad s_j^k = \left(\frac{1}{3}\right)^k \left(\frac{2}{3}\right)^k \quad k \geq 0, 1 \leq j \leq 2^k.$$

The self-similarity of this measure makes it useful in verifying the counterexample. But, it appears that the pair of weights  $(\omega, \sigma)$  *do not* satisfy the two weight inequality (1.1).

The construction of the other two example measures is closely related to the structure of the function  $H\omega$ . On each interval  $G_j^k$ ,  $H\omega$  is monotonically decreasing, from  $\infty$  at the left hand endpoint of  $G_j^k$ , to  $-\infty$  at the right hand endpoint. In particular,  $H\omega$  has a unique zero  $z_k^j$ . And this selection of points define  $\sigma$  as in (10.2), namely

$$\sigma = \sum_{k,j} s_j^k \delta_{z_k^j}.$$

Of course, we gain a substantial cancellation in the testing condition (1.4) by locating the point mass at the zero of  $H\omega$ .

We then define the third measure  $\ddot{\sigma}$  by taking  $z_j^k \in G_j^k$  to the unique point so that  $H\omega(z_j^k) = (3/2)^k$ . We then can easily check that the  $L^2$  inequality for  $(\omega, \sigma)$  does not hold:

$$\int |H\omega|^2 d\ddot{\sigma}(x) = \sum_{k=1}^{\infty} \sum_{j=1}^{2^k} \left(\frac{9}{4} \cdot \frac{2}{9}\right)^k = \infty.$$

The weight pair  $(\omega, \ddot{\sigma})$  can be seen to satisfy the  $A_2$  condition, the forward testing condition (1.3), but fail the backwards testing condition. Thus, this pair of weights provides an alternate example to those provided in [NaVo] and [NiTr]. We will not further discuss the measure  $\ddot{\sigma}$ .

We can calculate the rate at which  $H\omega$  blows up at the endpoints of the complementary intervals. The rate is a reflection of the fractal dimension of the Cantor set.

**Lemma 10.3.** *Write  $G_j^k = (a_j^k, b_j^k)$ . We have*

$$H\omega(a_j^k - c3^{-k}) \simeq (3/2)^k, \quad k \geq 1, 1 \leq j \leq 2^k,$$

*and a similar equality holds for  $b_j^k$ . (Implied constants can be taken absolute; signs will be reversed for  $b_j^k$ .)*

This in particular shows that the zeros  $z_j^k$  cannot move too far from the middle:

$$(10.4) \quad \sup_{j,k} \frac{|z_j^k - \tilde{z}_j^k|}{|G_j^k|} < \zeta < 1.$$

The points  $\tilde{z}_j^k$  satisfy a similar inequality. This indicates the sensitivity of the two weight inequality to the precise definition of the measures involved.

*Proof.* Fix  $k$ , and consider the numbers  $H\omega(a_j^k - c3^{-k})$  for  $1 \leq j \leq 2^k$ . These are monotonically increasing as the point of evaluation moves from left to right across the interval  $[0, 1]$ . So we should verify that

$$(10.5) \quad C_1(3/2)^k \leq H\omega(a_1^k + c3^{-k}) \leq H\omega(a_{2^k}^k + c3^{-k}) \leq C_2(3/2)^k$$

We consider the right hand inequality. Writing

$$\begin{aligned} H\omega(a_{2^k}^k + c3^{-k}) &= \int_{(G_{2^k}^k)^c} \frac{\omega(dy)}{a_{2^k}^k + c3^{-k} - y} \\ &\leq \int_0^{a_{2^k}^k} \frac{\omega(dy)}{a_{2^k}^k + c3^{-k} - y} \end{aligned}$$

Here, we have discarded that part of the domain of the integral where the integrand would be negative. Now, on the interval  $[0, a_{2^k}^k]$ , the support of  $\omega$  is contained in the set  $\bigcup_{\ell=1}^k I_{2^{\ell-1}}^\ell$ . Using this, we continue the estimate above as

$$\begin{aligned} H\omega(a_{2^k}^k + c3^{-k}) &\leq \sum_{\ell=1}^k \omega(I_{2^{\ell-1}}^\ell) \sup_{y \in I_{2^{\ell-1}}^\ell} \frac{1}{a_1^k + c3^{-k} - y} \\ &\lesssim c^{-1} \frac{2^{-k}}{3^{-k}} \sum_{\ell=1}^{k-1} \frac{2^{-\ell}}{3^{-\ell}} \lesssim c^{-1} (3/2)^k. \end{aligned}$$

It is useful to note for use below, that in this sum, the summand associated with  $k = \ell$  is the dominant one.

We consider the left hand inequality in (10.5). We split the support of  $\omega$  into the sets  $I_1^k, I_2^k, I_2^{k-1}, \dots, I_2^1$ . By the argument above, we have

$$\left| \sum_{\ell=1}^{k-1} H(\omega \mathbf{1}_{I_2^\ell})(a_1^k + c3^{-k}) \right| \leq A(3/2)^k,$$

where  $A$  is absolute, and we have yet to select  $c$ . Then, we have

$$H(\omega \mathbf{1}_{I_1^k \cup I_2^k}) = \int_{I_1^k} \frac{1}{a_j^k - c3^{-k} - y} - \frac{1}{a_j^k - (1+c)3^{-k} - y} \omega(dy)$$

$$\gtrsim c^{-1} 3^k \omega(I_1^k)$$

For  $0 < c < (2A)^{-1}$ , we conclude our Lemma.  $\square$

**10.2. The  $A_2$  Condition.** We verify that the usual  $A_2$  condition holds for the pair  $(\omega, \sigma)$ . Due to the property (10.4), this same argument will apply to the measures  $\dot{\sigma}$  and  $\ddot{\sigma}$ . The starting point is the estimate

$$\sigma(I_r^\ell) = \sum_{(k,j): z_j^k \in I_r^\ell} s_j^k = \sum_{k=\ell}^{\infty} 2^{k-\ell} \left(\frac{1}{3}\right)^k \left(\frac{2}{3}\right)^k = 2^{-\ell} \sum_{k=\ell}^{\infty} \left(\frac{2}{3}\right)^{2k} \approx 2^{-\ell} \left(\frac{2}{3}\right)^{2\ell} = s_r^\ell.$$

From this, it follows that we have

$$(10.6) \quad \frac{\sigma(I_j^k) \omega(I_j^k)}{|I_j^k|^2} \approx \frac{s_j^k \omega(I_j^k)}{|I_j^k|^2} = 1$$

The analogous condition for the Poisson or strengthened  $A_2$  condition also holds. Indeed, using the uniformity of  $\omega$ , one can verify

$$\begin{aligned} \mathbf{P}(I_r^\ell, \omega) &\lesssim \frac{\omega(I_r^\ell)}{|I_r^\ell|}, \\ \mathbf{P}(I_r^\ell, \sigma) &\lesssim \sum_{m=0}^{\infty} \frac{1}{2^m} \frac{\omega(I^{\ell+m})_\sigma}{\omega(I^{\ell+m})} \leq \sum_{m=0}^{\infty} \frac{1}{2^m} \frac{2^{-(\ell+m)} \left(\frac{2}{3}\right)^{2(\ell+m)}}{\left(\frac{1}{3}\right)^{\ell+m}} \lesssim \left(\frac{2}{3}\right)^{2\ell} \lesssim \frac{\sigma(I_r^\ell)}{|I_r^\ell|}. \end{aligned}$$

From this and (10.6), we see that

$$\mathbf{P}(I_r^\ell, \omega) \mathbf{P}(I_r^\ell, \sigma) \lesssim 1.$$

Let us consider an interval  $I \subset [0, 1]$ , and let  $k$  be the smallest integer such that  $z_j^k \in AI$ . Here  $A > 1$  is a large constant, dependent upon the constant in (10.4). We note that  $j$  is unique. For  $j < j'$ , it follows that for some  $j''$  we have  $z_{j'}^k < z_{j''}^{k-1} < z_{j'}^k$ . In particular, we will have  $\sigma(AI) \simeq \sigma(G_j^k)$ . Let us also assume that  $G_j^k \subset AI$ . Let  $I_r^{k-1} \supset G_j^k$ . It follows that we have

$$\mathbf{P}(I, \omega) \mathbf{P}(I, \sigma) \lesssim \mathbf{P}(I_r^{k-1}, \omega) \mathbf{P}(I_r^{k-1}, \sigma) \lesssim \frac{\sigma(I) \omega(I)}{|I| |I|} \lesssim 1.$$

The last case is  $G_j^k \not\supseteq AI$ . We then have

$$\mathbf{P}(I, \sigma) \simeq \frac{s_j^k |I|}{(|I| + \text{dist}(z_j^k, I))^2} \simeq \frac{s_j^k}{|I|}.$$

The last inequality follows from the definition of  $z_j^k$ , and fact that we must have  $\text{dist}(I, \partial G_j^k) > |I|$ , provided  $A$  is sufficiently large. We then have  $\mathbf{P}(I, \omega) \lesssim 2^{-k} \frac{|I|}{|G_j^k|}$ . And so we can estimate

$$\begin{aligned} \mathbf{P}(I, \sigma) \mathbf{P}(I, \omega) &\lesssim \frac{s_j^k}{|I|} 2^{-k} \frac{|I|}{|G_j^k|} \\ &\lesssim \frac{2^{-k} s_j^k}{|G_j^k|} \lesssim 1. \end{aligned}$$

**10.3. The Pivotal and Hybrid Conditions.** In this section, we show that the weight pair  $(\omega, \sigma)$  fails the dual Pivotal Condition, namely the Hybrid Condition with  $\epsilon = 0$  and the roles of  $\omega$  and  $\sigma$  reversed. But, they satisfy the Hybrid Condition for all  $0 < \epsilon \leq 2$ , and the dual Hybrid Condition for  $\epsilon_0 \leq \epsilon \leq 2$  for some  $\epsilon_0 < 2$ .

**10.3.1. Failure of the Pivotal Condition for  $\epsilon = 0$ .** Failure of the Pivotal Condition is straight forward. Indeed,  $I_1^\ell \subset I_1^{\ell-1} \subset \dots \subset I_1^0$  and so

$$\mathbf{P}(G_1^\ell, \omega) \approx \mathbf{P}(I_1^\ell, \omega) \approx \sum_{k=0}^{\ell} \frac{|I_1^\ell|}{|I_1^k|^2} \omega(I_r^k) \approx \sum_{k=0}^{\ell} \frac{3^{-\ell}}{3^{-2k}} 2^{-k} \approx \left(\frac{3}{2}\right)^\ell,$$

and similarly

$$\mathbf{P}(G_r^\ell, \omega) \approx \mathbf{P}(I_r^\ell, \omega) \approx \left(\frac{3}{2}\right)^\ell, \quad \text{all } r.$$

Considering the decomposition  $\dot{\bigcup}_{\ell,r} G_r^\ell \subset [0, 1]$  we thus have

$$\sum_{\ell,r} |G_r^\ell|_\sigma \mathbf{P}(G_r^\ell, \omega)^2 \approx \sum_{\ell=0}^{\infty} 2^\ell \left(\frac{1}{3}\right)^\ell \left(\frac{2}{3}\right)^\ell \left(\frac{3}{2}\right)^{2\ell} \approx \sum_{\ell=0}^{\infty} 1 = \infty,$$

which shows that the dual Pivotal Condition, the one dual to (1.7), fails.

**10.3.2. The dual Hybrid Condition for large  $\epsilon$ .** Next we show that the dual Hybrid Condition

$$\sum_{r=1}^{\infty} \sigma(I_r) \mathbf{E}(I_r, \sigma)^\epsilon \mathbf{P}(I_r, \mathbf{1}_{I_0} \omega)^2 \leq (\mathcal{E}_\epsilon^*)^2 \omega(I_0),$$

holds for all  $\epsilon_0 \leq \epsilon \leq 2$  where

$$\epsilon_0 = \frac{1}{\frac{\ln 3}{\ln 2} - \frac{1}{2}} \approx 0.92 < 2.$$

We need this estimate, which shows that with Energy, we can get an essential strengthening of the  $A_2$  condition.

**Proposition 10.7.** *For  $\epsilon \geq \epsilon_0$  and any interval  $I \subset [0, 1]$ , we have the inequality*

$$(10.8) \quad \sigma(I)E(I; \sigma)^\epsilon P(I; \omega)^2 \lesssim \omega(I).$$

*Proof.* We can assume that  $E(I; \sigma) \neq 0$ . Let  $k$  be the smallest integer for which there is a  $r$  with  $z_r^k \in I$ . And let  $n$  be the smallest integer so that for some  $s$  we have  $z_s^{k+n} \in I$  and  $z_s^{k+n} \neq z_r^k$ . We can estimate  $E(I; \sigma)$  in terms of  $n$ . Namely, we have

$$E(I; \sigma)^2 \lesssim \left(\frac{2}{9}\right)^n.$$

Indeed, the worst case is when  $s$  is not unique. Then, there are two choices of  $s$  – but not more. Let  $z_{s'}^{k+n} \in I$ , where  $s' \neq s$ . Then, note that we have

$$\frac{|I - \{z_r^k\}|_\sigma}{\sigma(I)} \simeq \left(\frac{2}{9}\right)^n.$$

and this leads to the estimate above, remembering the characterization of Energy as a variance term.

Next we note that  $\sigma(I) \approx \left(\frac{2}{9}\right)^k$ ,  $\omega(I) \geq 2^{-k-n}$ , and  $P(I; \omega) \simeq \left(\frac{3}{2}\right)^k$ . This specifies everything in (10.8), so we choose  $\epsilon$  so that

$$\left(\frac{2}{9}\right)^k \left(\frac{2}{9}\right)^{\frac{\epsilon}{2}n} \left(\frac{3}{2}\right)^{2k} \lesssim 2^{-k-n}.$$

This inequality will be true for all pairs of  $n, k$  if  $\epsilon_0 \leq \epsilon < 2$  where

$$\left(\frac{2}{9}\right)^{\frac{\epsilon_0}{2}} = \frac{1}{2}.$$

□

It is now clear that the pair of weights  $(\omega, \sigma)$  satisfy the dual Energy conditions  $\mathcal{E}_\epsilon^*$  for  $\epsilon_0 \leq \epsilon \leq 2$ . Let  $I_0 \subset [0, 1]$  and let  $\{I_r : r \geq 1\}$  be any partition of  $I_0$ . We appeal to (10.8) to see that

$$\sum_{r \geq 1} \sigma(I_r) E(I; \sigma)^\epsilon P(I; \omega)^2 \lesssim \sum_{r \geq 1} \omega(I_r) = \omega(I_0).$$

**10.3.3. The Hybrid Condition for positive  $\epsilon$ .** It remains to verify that the pair of measures  $(\omega, \sigma)$  satisfy the Hybrid Conditions for all  $0 \leq \epsilon \leq 2$ . We will establish the pivotal condition (1.7), i.e.  $\mathcal{E}_0 < \infty$ , which then implies that  $\mathcal{E}_\epsilon < \infty$  for all  $0 \leq \epsilon \leq 2$ . For this it suffices to show that the forward maximal inequality

$$(10.9) \quad \int M(f\sigma)^2 d\omega \leq C \int |f|^2 d\sigma$$

holds for the pair  $(\omega, \sigma)$ , and (10.9) in turn follows from the testing condition

$$(10.10) \quad \int M(\mathbf{1}_Q \sigma)^2 d\omega \leq C \int_Q d\sigma,$$

for all intervals  $Q$  (see [Saw1]). We will show (10.10) when  $Q = I_r^\ell$ , the remaining cases being an easy consequence of this one. For this we use the fact that

$$(10.11) \quad \mathcal{M}(\mathbf{1}_{I_r^\ell} \sigma)(x) \leq C \left(\frac{2}{3}\right)^\ell, \quad x \in E \cap I_r^\ell.$$

To see (10.11), note that for each  $x \in I_r^\ell$  that also lies in the Cantor set  $E$ , we have

$$\mathcal{M}(\mathbf{1}_{I_r^\ell} \sigma)(x) \leq \sup_{(k,j): x \in I_j^k} \frac{1}{|I_j^k|} \int_{I_j^k \cap I_r^\ell} d\sigma \approx \sup_{(k,j): x \in I_j^k} \frac{\left(\frac{1}{3}\right)^{k \vee \ell} \left(\frac{2}{3}\right)^{k \vee \ell}}{\left(\frac{1}{3}\right)^k} \approx \left(\frac{2}{3}\right)^\ell.$$

Now we consider for each fixed  $m$ , the approximations  $\omega^{(m)}$  and  $\sigma^{(m)}$  to the measures  $\omega$  and  $\sigma$  given by

$$\begin{aligned} d\omega^{(m)}(x) &= \sum_{i=1}^{2^m} 2^{-m} \frac{1}{|I_i^m|} \mathbf{1}_{I_i^m}(x) dx, \\ \sigma^{(m)} &= \sum_{k < m} \sum_{j=1}^{2^k} s_j^k \delta_{z_j^k}. \end{aligned}$$

For these approximations we have in the same way the estimate

$$\mathcal{M}(\mathbf{1}_{I_r^\ell} \sigma^{(m)})(x) \leq C \left(\frac{2}{3}\right)^\ell, \quad x \in \bigcup_{i=1}^{2^m} I_i^m.$$

Thus for each  $m \geq 1$  we have

$$\begin{aligned} \int_{I_r^\ell} \mathcal{M}(\mathbf{1}_{I_r^\ell} \sigma^{(m)})^2 d\omega^{(m)} &\leq C \sum_{i: I_i^m \subset I_r^\ell} \left(\frac{2}{3}\right)^{2\ell} 2^{-m} \\ &= C 2^{m-\ell} \left(\frac{2}{3}\right)^{2\ell} 2^{-m} = C s_r^\ell \approx C \int_{I_r^\ell} d\sigma. \end{aligned}$$

Taking the limit as  $m \rightarrow \infty$  yields the case  $Q = I_r^\ell$  of (10.10). This completes our proof of the pivotal condition, and hence also the Hybrid Conditions (1.10) for all  $0 \leq \epsilon \leq 2$ .

**10.4. The Testing Conditions.** As an initial step in verifying the forward testing condition (1.3) for the pair  $(\omega, \sigma)$ , we replace  $\sigma$  by the self-similar measure  $\dot{\sigma}$ , and exploit the self-similarity of both measures  $\omega$  and  $\dot{\sigma}$ :

$$\omega = \frac{1}{2} \text{Dil}_{\frac{1}{3}} \omega + \frac{1}{2} \text{Trans}_{\frac{2}{3}} \text{Dil}_{\frac{1}{3}} \omega \equiv \omega_1 + \omega_2,$$

$$\dot{\sigma} = \frac{2}{9} \text{Dil}_{\frac{1}{3}} \dot{\sigma} + \delta_{\frac{1}{2}} + \frac{2}{9} \text{Trans}_{\frac{2}{3}} \text{Dil}_{\frac{1}{3}} \dot{\sigma} \equiv \dot{\sigma}_1 + \delta_{\frac{1}{2}} + \dot{\sigma}_2.$$

We compute

$$\begin{aligned} \int |H\dot{\sigma}|^2 \omega &= \int \left| H \left( \dot{\sigma}_1 + \delta_{\frac{1}{2}} + \dot{\sigma}_2 \right) \right|^2 \omega_1 + \int \left| H \left( \dot{\sigma}_1 + \delta_{\frac{1}{2}} + \dot{\sigma}_2 \right) \right|^2 \omega_2 \\ &= (1 + \varepsilon) \left\{ \int |H\dot{\sigma}_1|^2 \omega_1 + \int |H\dot{\sigma}_2|^2 \omega_2 \right\} + \mathcal{R}_\varepsilon, \end{aligned}$$

where the remainder term  $\mathcal{R}_\varepsilon$  is easily seen to satisfy

$$\mathcal{R}_\varepsilon \lesssim_\varepsilon \mathcal{A}_2^2 \left( \int \dot{\sigma} \right),$$

since the supports of  $\delta_{\frac{1}{2}} + \dot{\sigma}_2$  and  $\omega_1$  are well separated, as are those of  $\delta_{\frac{1}{2}} + \dot{\sigma}_1$  and  $\omega_2$ . For this we first use  $(a + b)^2 \leq (1 + \varepsilon) a^2 + (1 + \frac{1}{\varepsilon}) b^2$  to obtain

$$\begin{aligned} &\int \left| H \left( \dot{\sigma}_1 + \delta_{\frac{1}{2}} + \dot{\sigma}_2 \right) \right|^2 \omega_1 \\ &\lesssim \int \left( |H(\dot{\sigma}_1)| + \left| H \left( \delta_{\frac{1}{2}} + \dot{\sigma}_2 \right) \right| \right)^2 \omega_1 \\ &\lesssim \int \left\{ (1 + \varepsilon) |H(\dot{\sigma}_1)|^2 + \left( 1 + \frac{1}{\varepsilon} \right) \left| H \left( \delta_{\frac{1}{2}} + \dot{\sigma}_2 \right) \right|^2 \right\} \omega_1, \end{aligned}$$

and then for example,

$$\begin{aligned} \int |H(\dot{\sigma}_2)|^2 \omega_1 &= \int_{[0, \frac{1}{3}]} \left| \int_{[\frac{2}{3}, 1]} \frac{1}{y - x} \dot{\sigma}(y) \right|^2 \omega(x) \\ &\lesssim \left[ \frac{1}{\frac{1}{3} |[0, 1]|} \right]^2 \dot{\sigma}([\frac{2}{3}, 1])^2 \omega([0, \frac{1}{3}]) \\ &\lesssim 9 \frac{\dot{\sigma}([0, 1]) \omega([0, 1])}{|[0, 1]|^2} \int \dot{\sigma}_2 \lesssim \mathcal{A}_2^2 \int \dot{\sigma}. \end{aligned}$$

But now we note that

$$\begin{aligned} \int |H\dot{\sigma}_1|^2 \omega_1 &= \frac{1}{2} \int |H\dot{\sigma}_1(x)|^2 \text{Dil}_{\frac{1}{3}} \omega(x) = \frac{1}{2} \int \left| H\dot{\sigma}_1 \left( \frac{x}{3} \right) \right|^2 \omega(x) \\ &= \frac{1}{2} \int \left| \int \frac{1}{z - \frac{x}{3}} \frac{2}{9} \text{Dil}_{\frac{1}{3}} \dot{\sigma}(z) \right|^2 \omega(x) = \frac{1}{2} \left( \frac{2}{9} \right)^2 \int \left| \int \frac{1}{\frac{z}{3} - \frac{x}{3}} \dot{\sigma}(z) \right|^2 \omega(x) \\ &= \frac{1}{2} \left( \frac{2}{9} \right)^2 9 \int |H\dot{\sigma}(x)|^2 \omega(x) = \frac{2}{9} \int |H\dot{\sigma}|^2 \omega, \end{aligned}$$

and similarly  $\int |H\dot{\sigma}_2|^2 \omega_2 = \frac{2}{9} \int |H\dot{\sigma}|^2 \omega$ . Thus we have

$$(10.15) \quad \int |H\dot{\sigma}|^2 \omega = \frac{2}{9} (1 + \varepsilon) \int |H\dot{\sigma}|^2 \omega + \frac{2}{9} (1 + \varepsilon) \int |H\dot{\sigma}|^2 \omega + \mathcal{R}_\varepsilon,$$

and provided  $\int |H\dot{\sigma}|^2 \omega$  is finite we conclude that

$$\int |H\dot{\sigma}|^2 \omega = \frac{1}{1 - \frac{4}{9}(1 + \varepsilon)} \mathcal{R}_\varepsilon \lesssim_\varepsilon \mathcal{A}_2^2 \left( \int \dot{\sigma} \right),$$

for  $\varepsilon > 0$  so small that  $1 - \frac{4}{9}(1 + \varepsilon) > 0$ .

To avoid making the assumption that  $\int |H\dot{\sigma}|^2 \omega$  is finite, we use approximations as follows. For each fixed  $m \geq 1$ , consider the approximations  $\omega^{(m)}$  and  $\dot{\sigma}^{(m)}$  to the measures  $\omega$  and  $\dot{\sigma}$  as in (10.12). We have the following self-similarity equations involving  $\omega^{(m)}$  and  $\dot{\sigma}^{(m)}$  that substitute for (10.13): for  $m \geq 2$ ,

$$\begin{aligned} \omega^{(m)} &= \frac{1}{2} \text{Dil}_{\frac{1}{3}} \omega^{(m-1)} + \frac{1}{2} \text{Trans}_{\frac{2}{3}} \text{Dil}_{\frac{1}{3}} \omega^{(m-1)} \equiv \omega_1^{(m)} + \omega_2^{(m)}, \\ \dot{\sigma}^{(m)} &= \frac{2}{9} \text{Dil}_{\frac{1}{3}} \dot{\sigma}^{(m-1)} + \delta_{\frac{1}{2}} + \frac{2}{9} \text{Trans}_{\frac{2}{3}} \text{Dil}_{\frac{1}{3}} \dot{\sigma}^{(m-1)} \equiv \dot{\sigma}_1^{(m)} + \delta_{\frac{1}{2}} + \dot{\sigma}_2^{(m)}. \end{aligned}$$

As above we compute that

$$\begin{aligned} \int |H\dot{\sigma}^{(m)}|^2 \omega^{(m)} &= \int \left| H \left( \dot{\sigma}_1^{(m)} + \delta_{\frac{1}{2}} + \dot{\sigma}_2^{(m)} \right) \right|^2 \omega_1^{(m)} + \int \left| H \left( \dot{\sigma}_1^{(m)} + \delta_{\frac{1}{2}} + \dot{\sigma}_2^{(m)} \right) \right|^2 \omega_2^{(m)} \\ &= (1 + \varepsilon) \left\{ \int |H\dot{\sigma}_1^{(m)}|^2 \omega_1^{(m)} + \int |H\dot{\sigma}_2^{(m)}|^2 \omega_2^{(m)} \right\} + \mathcal{R}_\varepsilon^{(m)}, \end{aligned}$$

where the remainder term  $\mathcal{R}_\varepsilon^{(m)}$  satisfies  $\mathcal{R}_\varepsilon^{(m)} \lesssim_\varepsilon \mathcal{A}_2^2(\int \dot{\sigma})$ , and also that

$$\begin{aligned} \int |H\dot{\sigma}_1^{(m)}|^2 \omega_1^{(m)} &= \frac{1}{2} \int |H\dot{\sigma}_1^{(m)}|^2 \text{Dil}_{\frac{1}{3}} \omega^{(m-1)}(x) = \frac{1}{2} \int \left| H\dot{\sigma}_1^{(m)} \left( \frac{x}{3} \right) \right|^2 \omega^{(m-1)}(x) \\ &= \frac{1}{2} \int \left| \int \frac{1}{z - \frac{x}{3}} \frac{2}{9} \text{Dil}_{\frac{1}{3}} \dot{\sigma}^{(m-1)}(z) \right|^2 \omega^{(m-1)}(x) \\ &= \frac{1}{2} \left( \frac{2}{9} \right)^2 \int \left| \int \frac{1}{\frac{z}{3} - \frac{x}{3}} \dot{\sigma}^{(m-1)}(z) \right|^2 \omega^{(m-1)}(x) \\ &= \frac{1}{2} \left( \frac{2}{9} \right)^2 9 \int |H\dot{\sigma}^{(m-1)}(x)|^2 \omega^{(m-1)}(x) = \frac{2}{9} \int |H\dot{\sigma}^{(m-1)}|^2 \omega^{(m-1)}, \end{aligned}$$

and  $\int |H\dot{\sigma}_2^{(m)}|^2 \omega_2^{(m)} = \frac{2}{9} \int |H\dot{\sigma}^{(m-1)}|^2 \omega^{(m-1)}$ . Thus we have

$$\int |H\dot{\sigma}^{(m)}|^2 \omega^{(m)} = \frac{4}{9} (1 + \varepsilon) \int |H\dot{\sigma}^{(m-1)}|^2 \omega^{(m-1)} + \mathcal{R}_\varepsilon^{(m)}, \quad m \geq 2.$$



Iterating this equality yields

$$\int |H\dot{\sigma}^{(m)}|^2 \omega^{(m)} = \left(\frac{4}{9}(1+\varepsilon)\right)^m \int |H\dot{\sigma}^{(0)}|^2 \omega^{(0)} + \sum_{j=0}^{m-1} \left(\frac{4}{9}(1+\varepsilon)\right)^j \mathcal{R}_\varepsilon^{(m-j)}, \quad m \geq 2,$$

from which we obtain

$$\int |H\dot{\sigma}^{(m)}|^2 \omega^{(m)} \lesssim_\varepsilon \mathcal{A}_2^2 \left( \int \dot{\sigma} \right), \quad m \geq 2,$$

with a constant  $C$  independent of  $m$ . Taking the limit as  $m \rightarrow \infty$  proves  $\int |H\dot{\sigma}|^2 \omega \leq C_\varepsilon \mathcal{A}_2^2 \left( \int \dot{\sigma} \right) < \infty$ .

This completes the proof of the forward testing condition (1.3) for the interval  $I = [0, 1]$ . The proof for the case  $I = I_j^k$  is similar using  $\mathcal{R}_\varepsilon(I_j^k) \leq C_\varepsilon \mathcal{A}_2^2 \left( \int_{I_j^k} \dot{\sigma} \right)$ , and the general case now follows without much extra work.

**Remark 10.16.** *The self-similarity argument above works on the forward testing condition because the central point mass  $\delta_{\frac{1}{2}}$  is a significant fraction  $\frac{5}{9}$  of the mass of  $\dot{\sigma}$  and is well separated from the measure  $\omega$  at all scales. This accounts for the fact that a mere fraction  $\frac{4}{9}$  of the left side of (10.15) appears on the right side. This argument fails to apply to the two weight inequality (10.5) itself because the measure  $f\dot{\sigma}$  need not have a significant proportion of its mass concentrated at the point  $\frac{1}{2}$ .*

Having verified the forward testing condition for the weight pair  $(\omega, \dot{\sigma})$ , we show that the forward testing condition (1.3) holds for  $(\omega, \sigma)$ . For this, we estimate the difference

$$\begin{aligned} \int_{I_r^\ell} |H\mathbf{1}_{I_r^\ell}(\sigma - \dot{\sigma})|^2 \omega &= \int_{I_r^\ell} \left| \sum_{(k,j): z_j^k \in I_r^\ell} s_j^k \left( \frac{1}{x - z_j^k} - \frac{1}{x - \dot{z}_k^j} \right) \right|^2 \omega(x) \\ &= \int_{I_r^\ell} \left| \sum_{(k,j): z_j^k \in I_r^\ell} s_j^k \left( \frac{z_j^k - \dot{z}_k^j}{(x - z_j^k)(x - \dot{z}_k^j)} \right) \right|^2 \omega(x) \\ &\lesssim C \int_{I_r^\ell} \left| \sum_{(k,j): z_j^k \in I_r^\ell} s_j^k \left( \frac{|I_j^k|}{|x - z_j^k|^2} \right) \right|^2 \omega(x). \end{aligned}$$

In the last line, we have used (10.4). Now for any fixed  $x$  in the support of  $\omega$  inside  $I_r^\ell$ , we have

$$\sum_{(k,j): z_j^k \in I_r^\ell} s_j^k \left( \frac{|I_j^k|}{|x - z_j^k|^2} \right) = \sum_{m=0}^{\infty} \sum_{(k,j): z_j^k \in I_r^\ell \text{ and } |x - z_j^k| \approx 3^{-m} |I_r^\ell|} s_j^k \left( \frac{|I_j^k|}{|x - z_j^k|^2} \right)$$

$$\begin{aligned}
&\lesssim \sum_{m=0}^{\infty} \sum_{(k,j): z_j^k \in I_r^\ell \text{ and } |x-z_j^k| \approx 3^{-m} |I_r^\ell|} \left(\frac{1}{3}\right)^k \left(\frac{2}{3}\right)^k \left(\frac{3^{-k}}{(3^{-m-\ell})^2}\right) \\
&\approx \sum_{m=0}^{\infty} \sum_{k \geq \ell+m} 2^{k-\ell-m} \left(\frac{1}{3}\right)^k \left(\frac{2}{3}\right)^k \left(\frac{3^{-k}}{(3^{-m-\ell})^2}\right) \\
&= \sum_{m=0}^{\infty} \sum_{k \geq \ell+m} \left(\frac{4}{27}\right)^k \left(\frac{9}{2}\right)^{(m+\ell)} \\
&= \sum_{m=0}^{\infty} \left(\frac{2}{3}\right)^{m+\ell} \approx \left(\frac{2}{3}\right)^\ell.
\end{aligned}$$

Thus we get

$$\int_{I_r^\ell} |H \mathbf{1}_{I_r^\ell}(\sigma - \dot{\sigma})|^2 \omega \lesssim \left(\frac{2}{3}\right)^{2\ell} \omega(I_r^\ell) = C^2 \left(\frac{2}{3}\right)^{2\ell} 2^{-\ell} \approx \sigma(I_r^\ell),$$

which yields

$$\begin{aligned}
\left(\int_{I_r^\ell} |H \mathbf{1}_{I_r^\ell} \sigma|^2 \omega\right)^{\frac{1}{2}} &\lesssim \left(\int_{I_r^\ell} |H \mathbf{1}_{I_r^\ell} \dot{\sigma}|^2 \omega\right)^{\frac{1}{2}} + \left(\int_{I_r^\ell} |H \mathbf{1}_{I_r^\ell}(\sigma - \dot{\sigma})|^2 \omega\right)^{\frac{1}{2}} \\
&\lesssim C \sqrt{\sigma(I_r^\ell)}
\end{aligned}$$

This is the case  $I = I_r^\ell$  of the forward testing condition (1.3) for the weight pair  $(\omega, \dot{\sigma})$ , and the general case follows easily from this.

Finally, we turn to the dual testing condition (1.4) for the weight pair  $(\omega, \sigma)$ . For interval  $I_r^\ell$  and  $z_j^k \in I_r^\ell$ , we claim that

$$(10.17) \quad |H(\mathbf{1}_{I_r^\ell} \omega)(z_j^k)| \lesssim P(I_r^\ell, \omega).$$

To see this let  $I_s^{\ell-1}$  denote the parent of  $I_r^\ell$  and  $I_{r+1}^\ell$  denote the other child of  $I_s^{\ell-1}$ . Then we have using  $H\omega(z_k^j) = 0$ ,

$$\begin{aligned}
H(\mathbf{1}_{I_r^\ell} \omega)(z_k^j) &= -H(\mathbf{1}_{(I_r^\ell)^c} \omega)(z_k^j) \\
&= -H(\mathbf{1}_{(I_s^{\ell-1})^c} \omega)(z_k^j) - H(\mathbf{1}_{I_{r+1}^\ell} \omega)(z_k^j).
\end{aligned}$$

Now we have using  $H(\omega)(z_r^\ell) = 0$  that

$$\begin{aligned}
H(\mathbf{1}_{(I_s^{\ell-1})^c} \omega)(z_k^j) &= H(\mathbf{1}_{(I_s^{\ell-1})^c} \omega)(z_r^\ell) - \left\{ H(\mathbf{1}_{(I_s^{\ell-1})^c} \omega)(z_r^\ell) - H(\mathbf{1}_{(I_s^{\ell-1})^c} \omega)(z_k^j) \right\} \\
&= -H(\mathbf{1}_{I_s^{\ell-1}} \omega)(z_r^\ell) - A,
\end{aligned}$$

where

$$A \equiv H(\mathbf{1}_{(I_s^{\ell-1})^c} \omega)(z_r^\ell) - H(\mathbf{1}_{(I_s^{\ell-1})^c} \omega)(z_k^j).$$

Combining equalities yields

$$H\left(\mathbf{1}_{I_r^\ell}\omega\right)\left(z_k^j\right)=H\left(\mathbf{1}_{I_s^{\ell-1}}\omega\right)\left(z_r^\ell\right)+A-H\left(\mathbf{1}_{I_{r+1}^\ell}\omega\right)\left(z_k^j\right).$$

We then have for  $(k, j)$  such that  $z_j^k \in I_r^\ell$ ,

$$\begin{aligned} \left|H\left(\mathbf{1}_{I_s^{\ell-1}}\omega\right)\left(z_r^\ell\right)\right| &\lesssim \frac{\omega(I_s^{\ell-1})}{|I_s^{\ell-1}|}, \\ |A| &\lesssim \int_{(I_s^{\ell-1})^c} \left|\frac{1}{x-z_r^\ell}-\frac{1}{x-z_j^k}\right| \omega(x) \lesssim \int_{(I_s^{\ell-1})^c} \frac{|I_r^\ell|}{|x-z_s^{\ell-1}|^2} \omega(x), \\ \left|H\left(\mathbf{1}_{I_{r+1}^\ell}\omega\right)\left(z_j^k\right)\right| &\lesssim \frac{\omega(I_s^{\ell-1})}{|I_s^{\ell-1}|}, \end{aligned}$$

which proves (10.17).

Now we compute using (10.17) and the estimate for  $P(I_r^\ell, \omega)$  above that

$$\begin{aligned} \int_{I_r^\ell} |H(\mathbf{1}_{I_r^\ell}\omega)|^2 d\sigma &= \sum_{(k,j): z_j^k \in I_r^\ell} |H(\mathbf{1}_{I_r^\ell}\omega)(z_j^k)|^2 s_j^k \leq C \sum_{(k,j): z_j^k \in I_r^\ell} |P(I_r^\ell, \omega)|^2 s_j^k \\ &\lesssim \sigma(I_r^\ell) \left(\frac{\omega(I_r^\ell)}{|I_r^\ell|}\right)^2 \lesssim \mathcal{A}_2 \omega(I_r^\ell). \end{aligned}$$

This is the case  $I = I_r^\ell$  of the dual testing condition (1.4) for the weight pair  $(\omega, \sigma)$ , and the general case follows easily from this.

## REFERENCES

- [BeMeSe] Y. BELOV, T. Y. MENGESTIE AND K. SEIP, *Unitary Discrete Hilbert transforms*, [arXiv:0911.0318v1](#) (2009).
- [BeMeSe2] Y. BELOV, T. Y. MENGESTIE AND K. SEIP, *Discrete Hilbert transforms on sparse sequences*, [arXiv:0912.2899v1](#) (2009).
- [CoSa] M. COTLAR AND C. SADOSKY, *A moment theory approach to the Riesz theorem on the conjugate functions with general measures*, *Studia Math.* **53** (1975), no. 1, 75-101.
- [CrMaPe] D. CRUZ-URIBE, J. M. MARTELL AND C. PÉREZ, *Sharp two weight inequalities for singular integrals, with applications to the Hilbert transform and the Sarason conjecture*, *Adv. Math.* **216** (2007), 647–676, MR{2351373}.
- [DaJo] DAVID, GUY, JOURNÉ, JEAN-LIN, *A boundedness criterion for generalized Calderón-Zygmund operators*, *Ann. of Math. (2)* **120** (1984), 371–397, MR763911 (85k:42041).
- [DrVo] DRAGIČEVIĆ, OLIVER AND VOLBERG, ALEXANDER, *Sharp estimate of the Ahlfors-Beurling operator via averaging martingale transforms*, *Michigan Math. J.* **51** (2003), no. 2, 415–435, MR1992955 (2004c:42030).
- [Hyt] HYTÖNEN, TUOMAS, *On Petermichl’s dyadic shift and the Hilbert transform*, *C. R. Math. Acad. Sci. Paris* **346** (2008), no. 21-22, 1133–1136, MR 2464252.
- [LaPeRe] LACEY, MICHAEL T., PETERMICHL, STEFANIE AND REGUERA, MARIA CARMEN, *Sharp  $A_2$  Inequality for Haar Shift Operators*, [arxiv0906.1941 \(su2009\)](#).
- [LaSaUr1] LACEY, MICHAEL T., SAWYER, ERIC T., URIARTE-TUERO, IGNACIO, *A characterization of two weight norm inequalities for maximal singular integrals*, [arxiv:0807.0246](#) (2008).

- [LaSaUr2] LACEY, MICHAEL T., SAWYER, ERIC T., URIARTE-TUERO, IGNACIO, *Two Weight Inequalities for Maximal Truncations of Dyadic Calderón-Zygmund Operators*, [arxiv:0911.3920](#) (2009).
- [LaSaUr3] LACEY, MICHAEL T., SAWYER, ERIC T., URIARTE-TUERO, IGNACIO, *Two Weight Inequalities for Discrete Positive Operators*, [arXiv:0911.3437](#) (2009).
- [Muc] MUCKENHOUT, BENJAMIN, *Hardy's inequality with weights*, Collection of articles honoring the completion by Antoni Zygmund of 50 years of scientific activity, I. *Studia Math.* **44** (1972), 31–38.
- [HuMuWh] R. HUNT, B. MUCKENHOUT AND R. L. WHEEDEN, *Weighted norm inequalities for the conjugate function and the Hilbert transform*, *Trans. Amer. Math. Soc.* **176** (1973), 227–251.
- [NaVo] F. NAZAROV AND A. VOLBERG, *The Bellman function, the two weight Hilbert transform, and the embeddings of the model space  $K_\theta$* , *J. d'Analyse Math.* **87** (2002), 385–414.
- [NTV1] F. NAZAROV, S. TREIL AND A. VOLBERG, *The Bellman function and two weight inequalities for Haar multipliers*, *J. Amer. Math. Soc.* **12** (1999), 909–928, MR{1685781 (2000k:42009)}.
- [NTV2] NAZAROV, F., TREIL, S. AND VOLBERG, A., *The Tb-theorem on non-homogeneous spaces*, *Acta Math.* **190** (2003), no. 2, MR 1998349 (2005d:30053).
- [NTV3] F. NAZAROV, S. TREIL AND A. VOLBERG, *Two weight inequalities for individual Haar multipliers and other well localized operators*, *Math. Res. Lett.* **15** (2008), no. 3, 583–597.
- [NTV4] F. NAZAROV, S. TREIL AND A. VOLBERG, *Two weight estimate for the Hilbert transform and corona decomposition for non-doubling measures*, preprint (2004) [arxiv:1003.1596](#)
- [NiTr] N. NIKOLSKI AND S. TREIL, *Linear resolvent growth of rank one perturbation of a unitary operator does not imply its similarity to a normal operator*, *J. Anal. Math.* **87** (2002), 415–431. MR1945291.
- [PeVoYu] F. PEHERSTORFER, A. VOLBERG AND P. YUDITSKII, *Two weight Hilbert transform and Lipschitz property of Jacobi matrices associated to hyperbolic polynomials*, *J. Funct. Anal.* **246** (2007), 1–30, MR{2316875}.
- [Pet1] S. PETERMICHL, *Dyadic shift and a logarithmic estimate for Hankel operators with matrix symbol*, *C. R. Acad. Sci. Paris* **330** (2000), 455–460, MR1756958 (2000m:42016).
- [Pet2] S. PETERMICHL, *The sharp bound for the Hilbert transform on weighted Lebesgue spaces in terms of the classical  $A_p$  characteristic*, *Amer. J. Math.* **129** no. 5 (2007), 1355–1375, MR2354322 (2008k:42066).
- [Pet3] PETERMICHL, STEFANIE, *The sharp weighted bound for the Riesz transforms*, *Proc. Amer. Math. Soc.* **136** no 4 (2008), 1237–1249, MR2367098 (2009c:42034).
- [PeVo] ST. PETERMICHL AND A. VOLBERG, *Why the Riesz transforms are averages of the dyadic shifts?*, preprint.
- [Saw1] E. SAWYER, *A characterization of a two-weight norm inequality for maximal operators*, *Studia Math.* **75** (1982), 1–11, MR{676801 (84i:42032)}.
- [Saw2] E. SAWYER, *A two weight weak type inequality for fractional integrals*, *Trans. A.M.S.* **281** (1984), 339–345, MR{719674 (85j:26010)}.
- [Saw3] E. SAWYER, *A characterization of two weight norm inequalities for fractional and Poisson integrals*, *Trans. A.M.S.* **308** (1988), 533–545, MR{930072 (89d:26009)}.
- [SaWh] E. SAWYER AND R. L. WHEEDEN, *Weighted inequalities for fractional integrals on Euclidean and homogeneous spaces*, *Amer. J. Math.* **114** (1992), 813–874.
- [Vol] A. VOLBERG, *Calderón-Zygmund capacities and operators on nonhomogeneous spaces*, CBMS Regional Conference Series in Mathematics (2003), MR{2019058 (2005c:42015)}.
- [Wit1] WITTWER, JANINE, *A sharp estimate on the norm of the martingale transform*, *Math. Res. Lett.* **7** no. 1 (2000), 1–12, MR1748283 (2001e:42022).
- [Wit2] WITTWER, JANINE, *A sharp estimate on the norm of the continuous square function*, *Proc. Amer. Math. Soc.* **130** no. 8 (2002), 2335–2342 (electronic), MR1897458 (2003j:42009).

- [Zhe] D. ZHENG, *The distribution function inequality and products of Toeplitz operators and Hankel operators*, J. Funct. Anal. **138** (1996), 477–501, MR{1395967 (97e:47040)}.

SCHOOL OF MATHEMATICS, GEORGIA INSTITUTE OF TECHNOLOGY, ATLANTA GA 30332  
*E-mail address:* lacey@math.gatech.edu

DEPARTMENT OF MATHEMATICS & STATISTICS, MCMASTER UNIVERSITY, 1280 MAIN STREET WEST,  
HAMILTON, ONTARIO, CANADA L8S 4K1  
*E-mail address:* sawyer@mcmaster.ca

DEPARTMENT OF MATHEMATICS, MICHIGAN STATE UNIVERSITY, EAST LANSING MI  
*E-mail address:* ignacio@math.msu.edu